Geometric kth shortest paths

Topi Talvitie

Master's thesis
UNIVERSITY OF HELSINKI
Department of Mathematics and Statistics

Helsinki, November 25, 2015

Tiedekunta — Fakultet — Faculty

Faculty of Science

Laitos — Institution — Department

Department of Mathematics and Statistics

Tekijä — Författare — Author

Topi Talvitie

Työn nimi — Arbetets titel — Title

Geometric kth shortest paths

Oppiaine — Läroämne — Subject

mathematics

Työn laji — Arbetets art — Level

Master's thesis

Aika — Datum — Month and year

November 25, 2015

Sivumäärä — Sidoantal — Number of pages

86

Tiivistelmä — Referat — Abstract

Finding shortest paths in planar domains bounded by polygons is a well-studied problem in computational geometry. However, in many applications, only finding the shortest path is not sufficient: we need to be able to generate a list of short paths among which we can choose the route. Simple detours to the shortest path are rarely better than the direct path, and therefore we should return paths that are essentially different. To ensure that, we limit our consideration to *locally shortest paths*, defined as the paths that cannot be made shorter by infinitesimal perturbations, or more intuitively, the paths that are "pulled taut" around the obstacles of the domain.

We use the first half of the thesis to present the definitions and the basic theory of locally shortest paths. We prove that they are always polygonal chains of certain type, and use this to describe a simple visibility graph based algorithm for finding the *kth shortest path*, i.e. the kth element in the list of locally shortest paths between given points ordered by length.

We prove that there is a unique way to change a locally shortest path continuously by moving its endpoints while keeping the path locally shortest. This result is used to show that the set of locally shortest paths with one fixed endpoint forms a covering space of the planar domain. We use this to to prove a connection between homotopy theory and locally shortest paths: each homotopy class contains exactly one locally shortest path, and that path is the shortest path in its homotopy class.

The covering space structure formed by locally shortest paths also gives rise to the idea of tracking the lengths of the locally shortest paths between a fixed point s and a point x, and drawing a map of the points x in which the order of lengths of the paths changes or the type of one of the locally shortest paths changes. The resulting map is the *kth shortest path map*, a subdivision of the domain into components such that the kth shortest path from s to any point within a single component is essentially the same. We analyze the structure and complexity of this map, concluding that we can use it for efficient queries of kth shortest paths from s to any point x.

Avainsanat — Nyckelord — Keywords

computational geometry, locally shortest paths, route planning

Säilytyspaikka — Förvaringsställe — Where deposited

Muita tietoja — Övriga uppgifter — Additional information

Contents

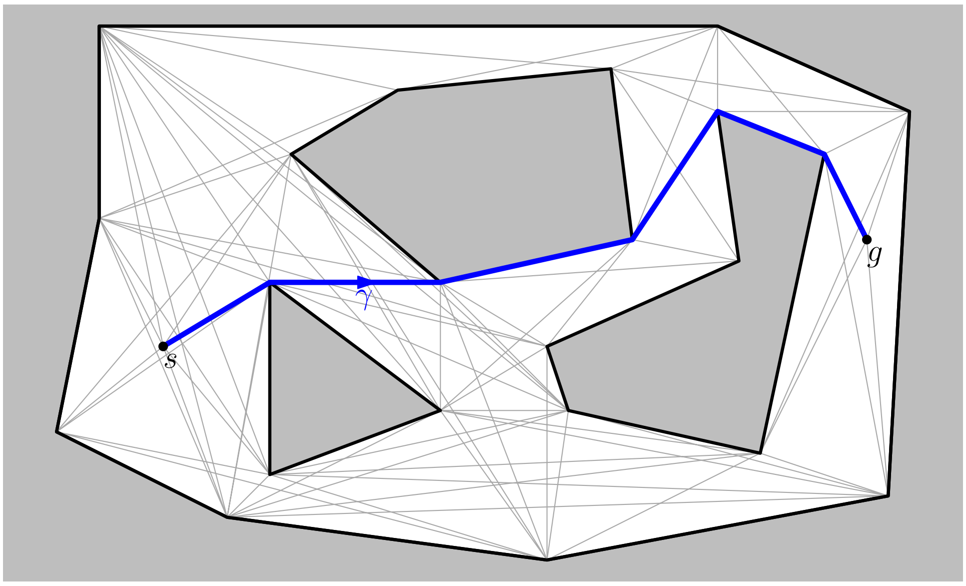


Figure 1.1: The visibility graph of a polygonal domain augmented with points s and g is drawn as gray lines. In the visibility graph, the pairs of vertices that can see each other are linked by an edge. The shortest path γ from s to g within the domain is the shortest path from s to g in the visibility graph.

---

1 Introduction

Route planning in planar domains is a practical problem with many applications. One of the simplest ways to model a planar environment is to use polygons, and therefore the problem of finding shortest paths in polygonal domains is an important and long-studied problem in computational geometry. A simple method to solve the problem is to reduce it to a shortest path problem in the visibility graph of the domain (Figure 1.1), which can then be solved using the Dijkstra algorithm. If n is the number of vertices and m is the number of edges in the visibility graph, then both the construction of the visibility graph and the Dijkstra algorithm run in $O(n \log n + m)$ time [5, 6, 12]. However, in the worst case m can be as large as $n(n-1)/2$, and therefore the methods based on visibility graphs will have at least quadratic worst-case time complexity.

Another, more geometric way to find shortest paths in polygonal domains is the *continuous Dijkstra* method: similarly to the Dijkstra algorithm, maintain the set of points closer than $r \geq 0$ to the source point s, and as r increases, the set will cover all the reachable points. In the graph version of the Dijkstra algorithm, this set is a discrete set of vertices, but in our case, it is an infinite set of points bounded by circular wavelets. The output of

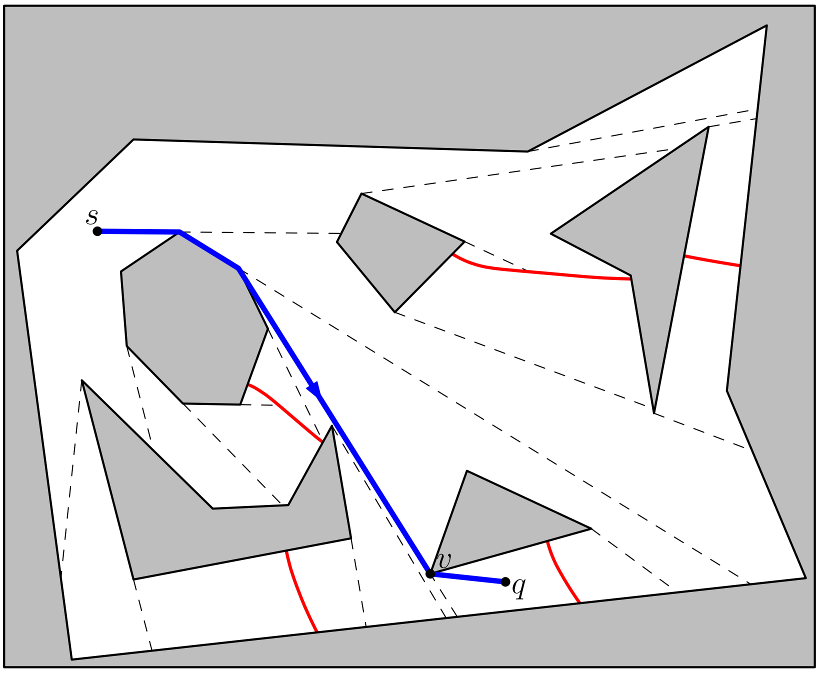


Figure 1.2: The shortest path map of point s, a subdivision of the domain into cells by hyperbolic segments (in red) and line segments (in dashed black), such that for all points q within one cell, the shortest path from s to q before the last segment vq is the same.

---

the continuous Dijkstra is the *shortest path map* (SPM) of s (Figure 1.2), a subdivision of the domain into components by the type of the shortest path. The edges of the subdivision consist of $O(n)$ line segments and segments of hyperbolic curves. We can build a standard point location query data structure for this subdivision in $O(n \log n)$ time to support querying the shortest path from s to any point in $O(\log n)$ time. An optimal algorithm that constructs this subdivision in $O(n \log n)$ time was developed in [10].

1.1 kth shortest paths

In many applications of geometric path planning, we are not interested only in the shortest path, but all paths that are reasonably short. For example, when choosing a flight path at a given flight level, we have a two-dimensional space in which we can fly freely. The path of the flight should be chosen so that it does not enter hazardous weather systems or no-fly zones. A simple way to solve the problem is to model the forbidden areas as polygons, and find the shortest path that does not pass through any polygon. However, despite its theoretical optimality, it could happen that this path is not feasible in practice for some reason. For example, the path could go through a too

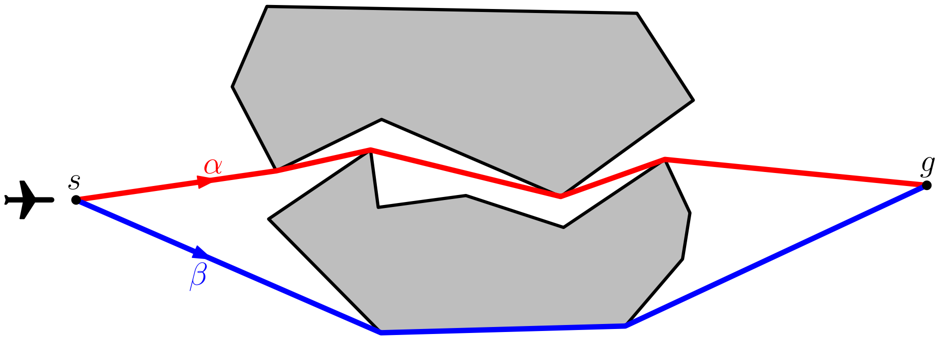


Figure 1.3: An example instance of the airplane routing problem in which we want to route a flight from s to g, avoiding hazardous weather systems represented as gray polygons. Even though α is the shortest feasible path, flying through a narrow channel between weather systems poses a high risk, and the pilot should choose a slightly longer but safer path β.

---

narrow channel between two weather systems (see Figure 1.3). Thus it is reasonable to algorithmically generate a set of short path options, and leave the final decision for the pilot. Similarly, in virtually every application of finding shortest paths, we have tradeoffs between the length of the path and other characteristics of the path such as the steepness of the turns, the visibility of the path from strategic locations or the general riskiness of the traversed areas. In practice, these characteristics are often hard to formulate or implement directly in the route planning algorithm, and thus we have to generate a set of short paths to be evaluated.

> *This thesis is devoted to defining geometric kth shortest paths, analyzing their structural properties and presenting algorithms for computing them.*

Finding the kth shortest path between given vertices in a graph is a long-studied problem. The current state-of-the-art algorithm finds the set of k shortest paths in $O(m + n \log n + k)$ time [3], where n is the number of vertices and m is the number of edges. The geometric equivalent of kth shortest paths was first studied in a recent paper [4]. However, the paper leaves many details only partly addressed, due to its limited length and large scope. We will derive a subset of the results using more mathematical rigor, deviating from the definitions of the paper in some cases for this purpose.

The exact definition of the problem in the geometric setting is not as intuitive as with graphs, because the set of possible path lengths is not discrete (it fills an entire interval of real numbers). We solve this by only considering *locally shortest paths*, i.e. paths that cannot be shortened by an

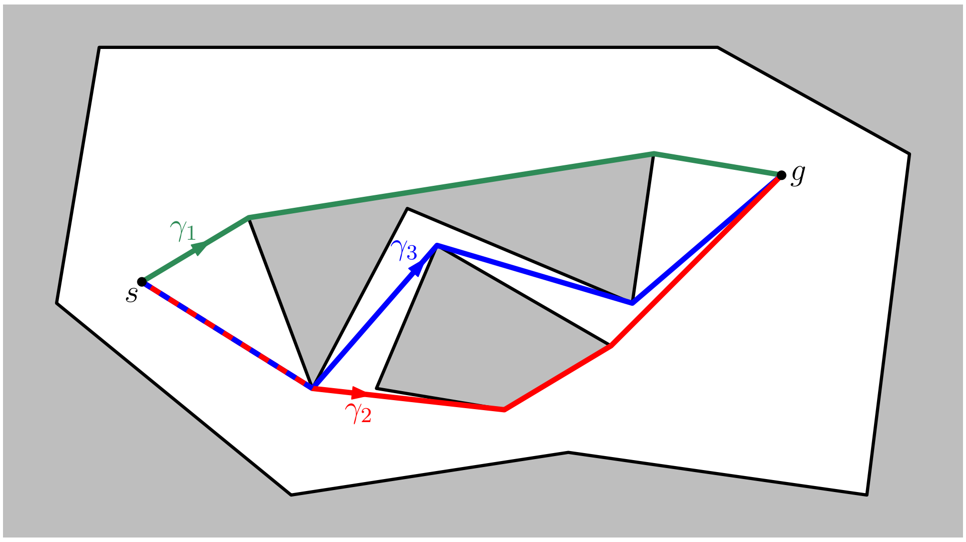


Figure 1.4: γ_1, γ_2 and γ_3 are the first three locally shortest paths from s to g in the domain when the paths are ordered by length. γ_1 is the first shortest path, γ_2 is the second shortest path and γ_3 is the third shortest path from s to g.

---

infinitesimal perturbation. Now the kth shortest path between given points is the kth element of the list of locally shortest paths ordered by length (see Figure 1.4 for an example). We dedicate the first half of the thesis to introducing the necessary theory for the definition of locally shortest paths and studying the covering space structure we get from the locally shortest paths by fixing one of the endpoints. We also prove that an equivalent definition for locally shortest paths is that they are exactly the shortest paths of their homotopy types (proving that our definition is equivalent to that of [4]).

After building the theory of locally shortest paths, we describe algorithms for finding the kth shortest path by reducing the problem into finding the kth shortest path in a modified visibility graph, and solving it with the algorithm of [3]. For the remainder of the thesis, we define and study the *kth shortest path map* (k-SPM), which is the higher-degree analogue of the SPM: a subdivision of the domain into cells such that one can query the kth shortest path from s to any point q simply by finding the cell that contains q. We study the structure of the k-SPM, filling in the details that were omitted in the paper [4], proving some additional results in the process. Similarly to the results of [10] for the 1-SPM, we get that the k-SPM is delimited by line segments and hyperbolic segments that do not intersect internally. We also prove a polynomial bound $O(k^3n^3)$ for the number of these segments. By using standard point location query algorithms [2, 11], we get that there exists a $O(k^3n^3(\log k + \log n))$-space structure that can be used to query the kth shortest path from s to any point in the domain in $O(\log k + \log n)$ time.

1.2 Overview

- Section 2: We give the basic definitions of polygonal domains, paths and lengths of paths, and prove some basic properties we will need later. These all are standard concepts whose definitions are folklore; we chose to state their exact definitions such that the objects are easy to work with in the following sections.

- Section 3: We develop the theory of locally shortest paths. In Subsection 3.1 we define the distance between paths as their Fréchet distance, which is a natural notion of distance for paths [1], and prove that it is a metric. We use it in Subsection 3.2 to define locally shortest paths and prove their basic properties. In Subsection 3.3, we prove that the set of locally shortest paths with one endpoint fixed forms a covering space of the domain, providing tools for considering local modifications to locally shortest paths. To the best of our knowledge, this is the first time locally shortest paths are defined like this. The paper [4] had intuition similar to our definition, but used homotopy for the exact definition, and proved things only on the intuitive level. A structure similar to our covering space was developed in [9] using triangulations.

- Section 4: We present a simple algorithm based on visibility graphs to compute k shortest paths between given points. We follow the idea given in [4], filling in the proofs and details about different query types and the output format of the paths.

- Section 5: We define the kth shortest path map, and study its structure. We analyze its boundary, decomposing it into $(k-1)$-walls, k-walls and k-windows, following the ideas of the paper [4] in which the structure was first described. However, because the presentation given in paper was very brief and skipped proper handling of some corner cases, we have to develop the exact definitions and proofs ourselves, adapting the ideas of the paper.

2 Preliminaries

In this section, we will define the basic terminology and the global setting the problem is posed in.

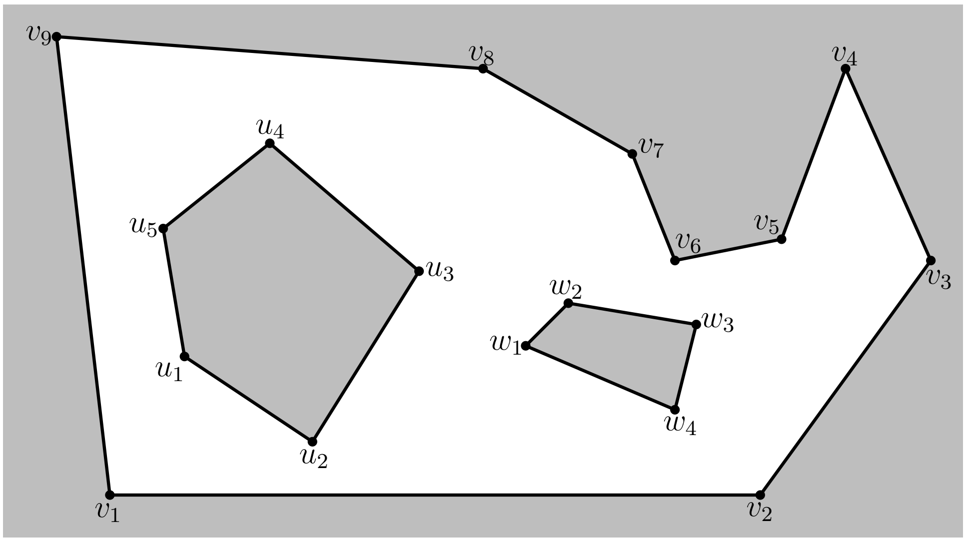


Figure 2.1: Example of a polygonal domain. The domain includes all the points in the white area and on the boundary drawn in black. The boundary consists of three components, all of which are polygons. The polygons have vertex lists $(u_1, u_2, \ldots, u_5)$, $(v_1, v_2, \ldots, v_9)$ and (w_1, w_2, w_3, w_4).

2.1 Polygonal domains

We work in the Euclidean plane, that is, the vector space $\mathbb{R}^2$ with the Euclidean norm defined as $(x, y) \mapsto \sqrt{x^2 + y^2} =: ||(x, y)||$. The distance between two points $a, b \in \mathbb{R}^2$ is the norm of their difference $||b - a||$.

We will consider only paths that are contained in a fixed subset of the plane (the walkable area). Our methods will require the domain to be a *polygonal domain*:

Definition 2.1. Set $X \subset \mathbb{R}^2$ is a *polygon* if there exists $n \geq 3$ and points $x_1, \ldots, x_n \in \mathbb{R}^2$ such that the edge sets $E_1, E_2, \ldots, E_n$ defined by

$$E_i = \{(1-t)x_{i-1} + tx_i \mid t \in [0, 1)\},$$

where we set $x_0 = x_n$, are pairwise disjoint and $X = \bigcup_{i=1}^{n} E_i$. Points $x_1, \ldots, x_n$ are called the *vertices* of the polygon.

Set $\Omega \subset \mathbb{R}^2$ is a *polygonal domain*, if

- $\partial\Omega$ has finitely many components, and every component is a polygon.
- Ω is closed, connected and bounded.

See Figure 2.1 for an example of how we will draw polygonal domains. The vertices of the component polygons of $\partial\Omega$ are also called the vertices of the polygonal domain Ω.

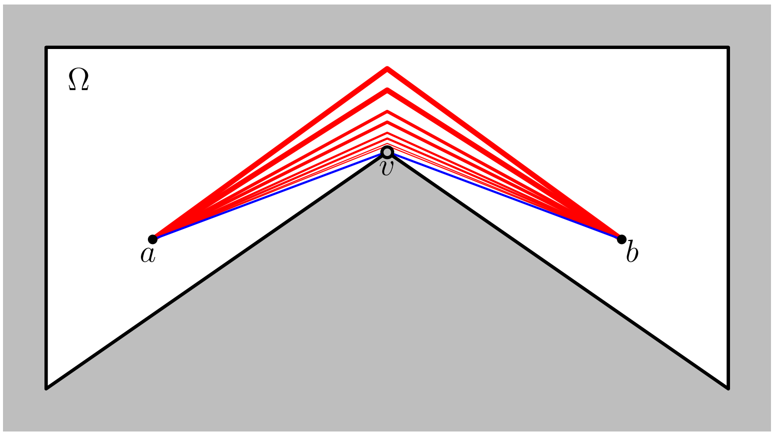


Figure 2.2: If the boundary point v is not contained in the domain Ω, then there is no shortest path from a to b. We cannot reach path length $||v-a|| + ||v-b||$ but we can get arbitrarily close to that.

The requirement that $\partial\Omega$ consists of finitely many polygons makes it possible to store $\partial\Omega$ using finite memory, since a polygon can be stored as a finite list of pairs of numbers and there is a finite number of those in $\partial\Omega$. This of course assumes that we can store a point in $\mathbb{R}^2$ as a pair of real numbers in constant amount of memory, which does not hold in actual computers. However, real numbers can be approximated by floating-point numbers such that in practice, the space and time requirements do not significantly affect the total complexity. Thus we work under the assumption that real numbers can be stored in constant space and arithmetic operations run in constant time. This model of computation is commonly known as *real RAM*.

The requirement that Ω is closed ensures that the boundary is contained in the domain. We need this, because otherwise there could be cases where there does not exist a shortest path in the domain between given points (see Figure 2.2). The requirements of closedness and boundedness also make Ω uniquely determined by $\partial\Omega$, because for every point outside $\partial\Omega$ we know that it is contained in the domain Ω if it is in the area bounded by an odd number of polygons of $\partial\Omega$. Thus because we can store $\partial\Omega$ in finite memory, we can also store Ω in finite memory.

The connectedness requirement of Ω is imposed only for convenience, as we can now always assume that paths between any two points exist. If we want to support disconnected domains, we can just process the domain one component at a time.

Now we will fix a polygonal domain and some special points we will work with. For algorithmic results, this will be in the input to the algorithm. For the sake of convenience, we will disallow the degenerate cases where three defining points are collinear. We could handle these degenerate cases as special cases in the theorems and algorithms, but as this is both trivial and tedious, we choose not to allow them.

Global definition 2.2. Let P be a polygonal domain and V be the set of its vertices. Let $s, g \in P$. We call s the *source point* and g the *target point*. We assume that $s, g \notin V$, and that no three points in $V \cup \{s, g\}$ are collinear.

The restriction that no three points of $V \cup \{s, g\}$ are collinear does not essentially limit the applicability of our methods, because if we view the configurations as tuples of $n = |V| + 2$ points in $\mathbb{R}^2$, so that one configuration is a point in the $2n$-dimensional space $\mathbb{R}^{2n}$, the degenerate cases where three points are collinear have measure zero, and thus any small random perturbation to the points have probability 1 of being non-degenerate.

2.2 Paths

Path and its length are natural concepts, and in the introduction we already used them in their intuitive meaning. In this subsection we will introduce a more exact definition of the set of paths with well-defined length. The end results of our algorithms will turn out to be polygonal chains. However, a priori we do not know that, and therefore most of our theoretical results handle a broader set of paths. The minimum requirement for paths is that they are continuous. Let us call such functions *curves*, and define some standard terminology for them.

Definition 2.3. A function $\gamma : [0, L] \to P$ for some $L \geq 0$ is a *curve*, if it is continuous. We say that γ is a curve from $\gamma(0)$ to $\gamma(L)$. Denote the set of curves from $a \in P$ to $b \in P$ by $\mathcal{C}_{ab}$, and the set of all curves $\bigcup_{a,b \in P} \mathcal{C}_{ab}$ by $\mathcal{C}$.

- Define the *subcurve* of range $[a, b] \subset [0, L]$ of the curve γ as the function $\gamma_{[a,b]} : [0, b - a] \to P$ given by

$$\gamma_{[a,b]}(t) = \gamma(t + a).$$

- For any $0 \leq x \leq L$, define the *x-prefix* of the curve γ as the function $\gamma_{[0,x]}$ denoted by $\gamma_{::x}$.

- For any $0 \leq x \leq L$, define the *x-suffix* of the curve γ as the function $\gamma_{[L-x,L]}$ denoted by $\gamma_{x::}$.

- Let $\alpha : [0, A] \to P$ and $\beta : [0, B] \to P$ be curves such that $\alpha(A) = \beta(0)$. Define the *concatenation* of α and β as the curve $\alpha\beta : [0, A + B] \to P$ given by

$$\alpha\beta(t) = \begin{cases} \alpha(t), & \text{if } t \leq A \\ \beta(t - A), & \text{if } t > A. \end{cases}$$

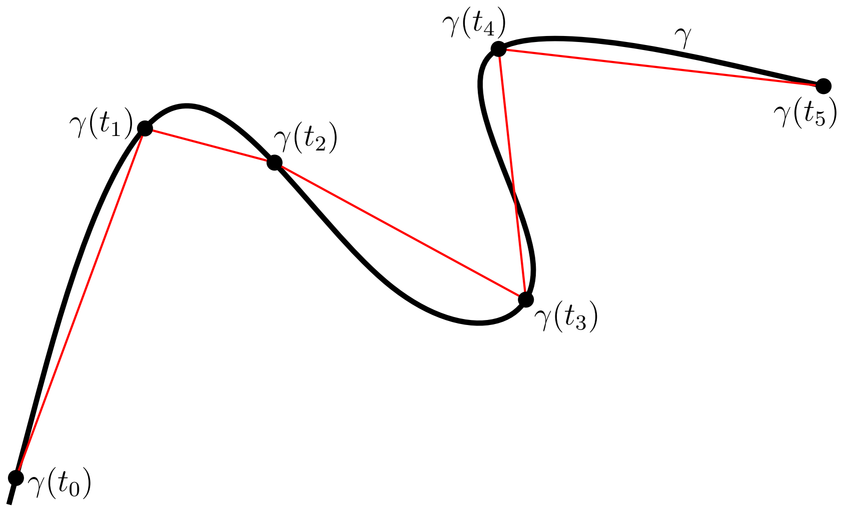


Figure 2.3: In Definition 2.4 we approximate curve $\gamma : [0, L] \to P$ by selecting finite sets of numbers in range $[0, L]$, and drawing polygonal lines through the corresponding points of γ. In the case of the figure, we select set $S = \{t_0, t_1, \ldots, t_5\}$, and get a lower estimate $A(S) = \sum_{k=1}^{5} ||\gamma(k) - \gamma(k-1)||$ for $|\gamma|$. As we choose more and more points, we get a closer estimate.

- Denote the reparameterization of γ from the unit interval by γ', defined as the curve $[0, 1] \to P$ given by
$$\gamma'(t) = \gamma(Lt).$$

However, it turns out that the set of curves is too broad for our use: there exists curves that have infinite length. Let us limit our consideration to the set of paths with finite length, called *rectifiable curves.*

Definition 2.4. Let $\gamma : [0, L] \to P$ be a curve. For any finite set $S \subset [0, L]$, we denote by $A(S)$ the length of the approximation of γ by a polygonal chain with vertices given by S, i.e.

$$A(S) = \sum_{k=1}^{|S|-1} ||\gamma(S_{[k+1]}) - \gamma(S_{[k]})||,$$

where by $S_{[k]}$ we denote the kth element of S for $1 \le k \le |S|$, i.e. $S_{[k]} \in S$ and $|S \cap (-\infty, S_{[k]}]| = k$. We define the length of a curve $\gamma : [0, L] \to P$ as the supremum over all such approximations:

$$|\gamma| = \sup_{S \subset [0,L],\ |S| < \infty} A(S),$$

We define that γ is a *rectifiable curve* if $|\gamma| < \infty$. See Figure 2.3 for an intuitive interpretation of this definition.

We could just use the set of rectifiable curves as the set of paths we work with, but the problem is that in that case, paths $\alpha, \beta : [0, 1] \to P$ defined by $\alpha(t) = (t, 0)$ and $\beta(t) = (t^2, 0)$ would be considered distinct paths, even though both paths traverse the segment from $(0, 0)$ to $(1, 0)$, and intuitively would seem to be the same path. The only difference between these curves is the speed they advance in. To remove this ambiguity, our definition will require all paths to advance in speed 1, like α.

Definition 2.5. A rectifiable curve $\gamma : [0, L] \to P$ is a *path* if for all intervals $[t_1, t_2] \subset [0, L]$,

$$|\gamma_{[t_1,t_2]}| = t_2 - t_1.$$

Denote the set of paths from $a \in P$ to $b \in P$ (that is, the set of paths in $\mathcal{C}_{ab}$) by $\mathcal{P}_{ab}$, and the set of all paths by $\mathcal{P}$.

This definition ensures that the length of the path between $\gamma(a)$ and $\gamma(b)$ is given by the difference of the parameters $b - a$. In some sources the term used is *naturally parameterized.* It is not necessarily clear why this restriction does not rule out many paths, since a priori there might be some paths that we cannot naturally parameterize. We will prove that this is not the case in Theorem 2.9, but for that we need some basic properties of subcurve lengths shown in the following lemmas.

Lemma 2.6. *Let $\gamma : [0, L] \to P$ be a curve. Define $A(S)$ for all finite $S \subset [0, L]$ similarly to Definition 2.4. Now adding a point $t \in [0, L] \setminus S$ to S does not decrease the approximation value $A(S)$, i.e. $A(S \cup \{t\}) \geq A(S)$.*

Proof. If $t < \min S$, then

$$A(S \cup \{t\}) - A(S) = ||\gamma(\min S) - \gamma(t)|| \geq 0.$$

If $t > \max S$, then

$$A(S \cup \{t\}) - A(S) = ||\gamma(t) - \gamma(\max S)|| \geq 0.$$

The remaining case is that $\min S < t < \max S$. Now if $x = \max(S \cap [0, b])$ and $y = \min(S \cap [b, L])$,

$$A(S \cup \{t\}) - A(S) = ||\gamma(b) - \gamma(x)|| + ||\gamma(y) - \gamma(b)|| - ||\gamma(y) - \gamma(x)||,$$

which is nonnegative by the triangle inequality. $\square$

Lemma 2.7. *Let $\gamma : [0, L] \to P$ be a curve, and $0 \leq a \leq b \leq c \leq L$. Now we can compute subcurve length by parts as follows:*

$$|\gamma_{[a,c]}| = |\gamma_{[a,b]}| + |\gamma_{[b,c]}|.$$

This holds also in cases where γ is not a rectifiable curve, and some subcurve lengths might be infinite.

Proof. By Lemma 2.6 we may restrict the approximation sets in the definition of $|\gamma_{[a,c]}|$ to always contain b:

$$|\gamma_{[a,c]}| = \sup_{\substack{b\in S\subset[a,c]\\|S|<\infty}} A(S),$$

Now, splitting $A(S)$ into two parts we get that

$$|\gamma_{[a,c]}| = \sup_{\substack{b\in S\subset[a,c]\\|S|<\infty}} (A(S\cap[0,b]) + A(S\cap[b,L])).$$

As the selection of $S\cap[0,b]$ and $S\cap[b,L]$ are completely independent, we can move the sum out of the supremum to get

$$|\gamma_{[a,c]}| = \sup_{\substack{b\in S_1\subset[a,b]\\|S_1|<\infty}} A(S_1) + \sup_{\substack{b\in S_2\subset[b,c]\\|S_2|<\infty}} A(S_2) = |\gamma_{[a,b]}| + |\gamma_{[b,c]}|.$$

□

Lemma 2.8. *Let $\gamma : [0,L] \to P$ be a rectifiable curve. The prefix length function $f : [0,L] \to [0,|\gamma|]$ of γ defined by $f(t) = |\gamma_{[0,t]}|$ is a non-decreasing continuous surjection.*

Proof. f is non-decreasing, because if $0 \le a \le b \le L$, then $[0,a] \subset [0,b]$, so the supremum over finite sets $S \subset [0,a]$ in Definition 2.4 is at most the supremum over same function over finite sets $S \subset [0,b]$. Clearly $f(0) = 0$ and by definition we know that $f(L) = |\gamma|$. We only need to be prove that f is continuous, since now surjectivity follows from that.

Let $\epsilon > 0$. Because now $|\gamma| = \sup_{S\subset[0,L],|S|<\infty} A(S)$, we can select $S_\epsilon \subset [0,L]$ with $|S_\epsilon| < \infty$ such that

$$A(S_\epsilon) > |\gamma| - \epsilon.$$

By Lemma 2.6 we may assume that 0 and L are in the set S_ϵ. Let $t \in [0,L]$. Define sets $S'_\epsilon = (S_\epsilon \cap [0,t]) \cup \{t\}$ and $S''_\epsilon = (S_\epsilon \cap [t,L]) \cup \{t\}$. Now

$$A(S'_\epsilon) + A(S''_\epsilon) = A(S_\epsilon \cup \{t\}) \ge A(S_\epsilon).$$

By the definition of subcurve length, we know that $A(S'_\epsilon) \le |\gamma_{[0,t]}|$ and $A(S''_\epsilon) \le |\gamma_{[t,L]}|$. Using these inequalities we obtain also a lower bound for $A(S'_\epsilon)$:

$$A(S'_\epsilon) \ge A(S_\epsilon) - A(S''_\epsilon) > |\gamma| - \epsilon - |\gamma_{[t,L]}|.$$

From Lemma 2.7 we get that $|\gamma_{[0,t]}| + |\gamma_{[t,L]}| = |\gamma|$, which yields that

$$A(S'_\epsilon) > |\gamma|_{[0,t]} - \epsilon.$$

Define for all $k \in \mathbb{Z}_+$ function $g_k : [0, L] \to \mathbb{R}$ by

$$g_k(t) = A((S_{1/k} \cap [0, t]) \cup \{t\}).$$

Now g_k is continuous, because for all consecutive pairs of elements a, b in the set $S_{1/k}$, if $t \in [a, b]$ then

$$g_k(t) - g_k(a) = ||\gamma(t) - \gamma(a)||.$$

From the inequalities we proved earlier, we now know that $|g_k(t) - f(t)| < 1/k$ for all $t \in [0, L]$. Thus the function sequence $(g_k)_{k=1}^{\infty}$ converges uniformly to f, and because g_k is continuous for all $k \in \mathbb{Z}_+$, f is continuous. $\square$

Theorem 2.9. *If $\gamma : [0, L] \to P$ is a rectifiable curve, then there exists a path $\alpha : [0, |\gamma|] \to P$ such γ is a reparameterization of α, i.e. there exists a nondecreasing surjection $\phi : [0, |\gamma|] \to [0, L]$ such that $\gamma = \alpha \circ \phi$.*

Proof. Define ϕ as the length function from Lemma 2.8, i.e.

$$\phi(t) = |\gamma_{[0,t]}|.$$

Now by Lemma 2.8, ϕ is a non-decreasing continuous surjection. Define function $\xi : [0, |\gamma|] \to [0, L]$ by

$$\xi(x) = \min \phi^{-1}\{x\}.$$

This function is well-defined, because by the surjectivity and continuity of ϕ we know that for all $x \in [0, |\gamma|]$ it holds that $\phi^{-1}\{x\}$ is nonempty and closed, and thus the minimum exists. We also get that $\phi \circ \xi(x) = x$ for all $x \in [0, |\gamma|]$ directly from the definition. Because ϕ is non-decreasing, ξ is also non-decreasing.

Define $\alpha : [0, |\gamma|] \to P$ by $\alpha = \gamma \circ \xi$. Let $t \in [0, L]$. Define $t' = \xi \circ \phi(t)$. Now

$$\phi(t') = \phi(\xi \circ \phi(t)) = \phi \circ \xi(\phi(t)) = \phi(t),$$

and therefore $|\gamma|_{[0,t']} = |\gamma|_{[0,t]}$. By the minimality of $t' = \xi \circ \phi(t)$ in $\phi^{-1}\{\phi(t)\}$, we know that $0 \le t' \le t$. Now using Lemma 2.7 we can write $|\gamma|_{[0,t]}$ as $|\gamma|_{[0,t']} + |\gamma|_{[t',t]}$, thus yielding

$$|\gamma|_{[t',t]} = |\gamma|_{[0,t]} - |\gamma|_{[0,t']} = \phi(t) - \phi(t') = 0.$$

By choosing $S = \{t', t\}$ in Definition 2.4 we get a lower bound $||\gamma(t') - \gamma(t)||$ for $|\gamma|_{[t',t]} = 0$. Thus $\gamma(t) = \gamma(t') = \gamma(\xi \circ \phi(t)) = \gamma \circ \xi(\phi(t)) = \alpha \circ \phi(t)$. Because this holds for all $t \in [0, L]$, $\gamma = \alpha \circ \phi$.

Now the only thing remaining is to prove that α is a path. It suffices to prove that $|\alpha|_{[0,x]} = x$ for all $x \in [0, |\gamma|]$, because then by Lemma 2.7 we obtain that for any interval $[a, b] \subset [0, |\gamma|]$,

$$|\alpha|_{[a,b]} = |\alpha|_{[0,b]} - |\alpha|_{[0,a]} = b - a.$$

Let $x \in [0, |\gamma|]$. We will first prove that $|\alpha|_{[0,x]} = |\gamma|_{[0,\xi(x)]}$. By the definition of curve length,

$$|\gamma|_{[0,\xi(x)]} = \sup_{\substack{S \subset [0,\xi(x)] \\ |S| < \infty}} \sum_{k=1}^{|S|-1} ||\gamma(S_{[k+1]}) - \gamma(S_{[k]})||,$$

$$|\alpha|_{[0,x]} = \sup_{\substack{S' \subset [0,x] \\ |S'| < \infty}} \sum_{k=1}^{|S'|-1} ||\alpha(S'_{[k+1]}) - \alpha(S'_{[k]})||.$$

Now for any finite $S \subset [0, \xi(x)]$, by setting $S' = \phi S$ we get that

$$\sum_{k=1}^{|S'|-1} ||\alpha(S'_{[k+1]}) - \alpha(S'_{[k]})|| = \sum_{k=1}^{|S|-1} ||\gamma(S_{[k+1]}) - \gamma(S_{[k]})||,$$

because ϕ is nondecreasing and $\gamma = \alpha \circ \xi$. Thus $|\gamma|_{[0,\xi(x)]} \leq |\alpha|_{[0,x]}$. Similarly for any finite $S' \subset [0, x]$, by setting $S = \xi S'$ we get the same equation, because $\alpha = \gamma \circ \xi$, which now yields $|\gamma|_{[0,\xi(x)]} \geq |\alpha|_{[0,x]}$. Combining the results, we get that $|\alpha|_{[0,x]} = |\gamma|_{[0,\xi(x)]}$.

By the definition of ξ, $\phi(\xi(x)) = x$, which by the definition of ϕ means that $|\gamma|_{[0,\xi(x)]} = x$. Combining this with the result $|\gamma|_{[0,\xi(x)]} = |\alpha|_{[0,x]}$ yields the result $|\alpha|_{[0,x]} = x$. Therefore α is a path. □

Now, let us introduce some basic properties and notation for this set of paths we defined.

Theorem 2.10. *All subcurves, prefixes, suffixes and concatenations (as defined in 2.3) of paths are also paths. We call a subcurve of a path a* subpath*.*

Proof.
Subpath. *If γ is a path, then $\gamma_{[a,b]}$ is a path for all $[a, b] \subset [0, |\gamma|]$.*
Let $[x, y] \subset [0, b - a]$. By the definition of subcurves, for all $t \in [0, y - x]$,

$$\gamma_{[a,b][x,y]}(t) = \gamma(a + x + t),$$

so by applying the definition curve lengths, we get that

$$|\gamma_{[a,b][x,y]}| = \sup_{\substack{S \subset [0,y-x] \\ |S| < \infty}} \sum_{k=1}^{|S|-1} ||\gamma_{[a,b][x,y]}(S_{[k+1]}) - \gamma_{[a,b][x,y]}(S_{[k]})||$$

$$= \sup_{\substack{S \subset [a+x,a+y] \\ |S| < \infty}} \sum_{k=1}^{|S|-1} ||\gamma(S_{[k+1]}) - \gamma(S_{[k]})|| = |\gamma_{[a+x,a+y]}| = y - x,$$

which proves the claim.

Concatenation. *If $\alpha \in \mathcal{P}_{ab}$ and $\beta \in \mathcal{P}_{bc}$ for some $a, b, c \in P$, then the concatenation $\gamma = \alpha\beta$ is a path.*
Let $[a, b] \subset [0, |\alpha| + |\beta|]$. If $[a, b] \subset [0, |\alpha|]$, then $|\gamma_{[a,b]}| = |\alpha_{[a,b]}|$, and if $[a, b] \subset [|\alpha|, |\alpha| + |\beta|]$, by translating the coordinates by $|\alpha|$ we get that $|\gamma_{[a,b]}| = |\beta_{[a-|\alpha|,b-|\alpha|]}|$. Because α and β are paths, in both cases we get that $|\gamma|_{[a,b]} = b - a$.

Otherwise $a < |\alpha|$ and $b > |\alpha|$. In that case, we can compute the subcurve length in parts by Lemma 2.7, and reduce to the previous cases

$$|\gamma_{[a,b]}| = |\gamma_{[a,|\alpha|]}| + |\gamma_{[|\alpha|,b]}| = (|\alpha| - a) + (b - |\alpha|) = b - a.$$

Thus γ is a path. □

Definition 2.11.

- The *constant path* at $x \in P$, defined as the constant curve function $\alpha : \{0\} \to P$ with $\alpha(0) = x$, is denoted by $[x]$.
- The *segment path* from $a \in P$ to $b \in P$ is the affine curve function $\beta : [0, ||b - a||]$ defined by

$$\beta(t) = a + \frac{t}{||b - a||}(b - a)$$

for all $t \in [0, ||b - a||]$ if $a \neq b$, and the constant path $[a]$ if $a = b$. It is denoted by $[a, b]$.

- The *polygonal chain* γ through points $p_0, p_1, \ldots, p_n$ defined as the concatenation $[p_0, p_1][p_1, p_2] \cdots [p_{n-1}, p_n]$ is denoted by $[p_0, p_1, \ldots, p_n]$.

Proof that the defined curves are paths.

- α is a path, because $\alpha_{[0,0]} = 0$.
- β is a path, because for any $S \subset [x, y] \subset [0, ||b - a||]$,

$$\begin{aligned}\sum_{k=1}^{|S|-1} ||\beta(S_{[k+1]} - S_{[k]})|| &= \sum_{k=1}^{|S|-1} \left|\left|\frac{S_{[k+1]} - S_{[k]}}{||b - a||}(b - a)\right|\right| \\ &= \sum_{k=1}^{|S|-1} |S_{[k+1]} - S_{[k]}| \\ &= \max S - \min S \leq y - x,\end{aligned}$$

and value $y - x$ is achieved by setting $S = \{x, y\}$.

- γ is a path, because it is a concatenation of segment paths with matching endpoints.

$\square$

We also easily get the intuitive result that a path from a to b cannot have length less than $||b - a||$, and the length $||b - a||$ can only be achieved by a segment path.

Theorem 2.12. *Let $a, b \in P$ and $\gamma \in \mathcal{P}_{ab}$. Then*

- *γ is 1-Lipschitz, i.e. for all $x, y \in [0, |\gamma|]$,*

$$||\gamma(x) - \gamma(y)|| \leq |x - y|.$$

- *$|\gamma| \geq ||b - a||$.*
- *$|\gamma| = ||b - a||$ if and only if $\gamma = [a, b]$.*

Proof. The 1-Lipschitz property of γ follows directly by choosing $S = \{x, y\}$ in the supremum of Definition 2.4, yielding

$$||\gamma(x) - \gamma(y)|| \leq |x - y|$$

for all $x, y \in [0, |\gamma|]$. If we choose $x = 0$ and $y = |\gamma|$, we get the second result $|\gamma| \geq ||b - a||$. If $\gamma = [a, b]$, we get $|\gamma| = ||b - a||$ directly from Definition 2.11.

Assume that $|\gamma| = ||b - a||$. Let $t \in [0, |\gamma|]$. Denote $x = \gamma(t)$. Now if we define $\alpha = \gamma_{[0,t]}$ and $\beta = \gamma_{[t,\gamma]}$, $|\alpha| + |\beta| = ||b - a||$. From the first result of this theorem we get that now $||x - a|| \leq |\alpha|$ and $||x - b|| \leq |\beta|$, and from the triangle inequality in Euclidean norm $|| \cdot ||$ we get that

$$||x - a|| + ||x - b|| \geq ||b - a|| = |\alpha| + |\beta|.$$

Now $||x - a|| = |\alpha|$ and $||x - b|| = |\beta|$, because otherwise the above inequality does not hold. Now equality holds in the triangle inequality, which implies that x is on the segment from a to b. The only point on that segment on distance $|\alpha| = t$ from a is $[a, b](t)$. Thus $\gamma = [a, b]$. $\square$

3 Locally shortest paths

In this section, we will define *locally shortest paths*, which will be used to define kth shortest paths in polygonal domains. We do this by first defining a metric space for paths such that the neighborhoods of a path consist of small deviations to the path, and then defining locally shortest paths as local minimums of the path length function. Our definition differs from the method used in [4], where locally shortest paths are defined as the shortest paths in their homotopy type. However, we will prove in subsection 3.4 that our definition is equivalent with this definition.

3.1 Path distance

The distance between paths does not have an immediate natural definition. If two paths have equal lengths, then the natural definition is to use some norm of the difference of the paths, but the case when the paths do not have equal lengths is more problematic. We choose to define path length as a process where we traverse both paths from start to finish, but at varying speeds, and the distance is the infimum of the maximum distances over all these kinds of traversals. This kind of distance is called the *Fréchet distance* [1].

Definition 3.1. Let $\alpha, \beta \in \mathcal{P}$. Define the *distance* d between α and β by

$$d(\alpha, \beta) = \inf_{u,v \in R} \max_{t \in [0,1]} ||\beta'(v(t)) - \alpha'(u(t))||,$$

where α' and β' are the reparameterizations of α and β from the unit interval as defined in Definition 2.3, and R is the set of nondecreasing surjections $[0, 1] \to [0, 1]$.

Well-definedness proof. Because curves α' and β' are continuous, and any $f \in R$ is continuous, it holds that $\beta \circ v - \alpha \circ u$ is continuous. Therefore the image of compact set $[0, 1]$ is compact, and the maximum exists. □

Sometimes the following alternative definition for the path distance is easier to use.

Lemma 3.2. *Let F be the set of increasing bijections $[0, 1] \to [0, 1]$. Now if $\alpha, \beta \in \mathcal{P}$, then*

$$d(\alpha, \beta) = \inf_{f \in F} \max_{t \in [0,1]} ||\beta'(f(t)) - \alpha'(t)||.$$

Proof. The maximum exists again, because the function $\beta' \circ f - \alpha'$ is continuous. Denote

$$d'(\alpha, \beta) = \inf_{f \in F} \max_{t \in [0,1]} ||\beta'(f(t)) - \alpha'(t)||.$$

Now we want to prove that $d' = d$.

It holds that $d' \geq d$, because if $f \in F$, by choosing $u = \mathrm{id}_{[0,1]}$ and $v = f$ we get that $u, v \in R$ and

$$\max_{t \in [0,1]} ||\beta'(f(t)) - \alpha'(t)|| = \max_{t \in [0,1]} ||\beta'(v(t)) - \alpha'(u(t))||.$$

Now it suffices to prove that $d' \leq d$. Let $\alpha, \beta \in \mathcal{P}$ and $\epsilon > 0$. Select $u, v \in R$ such that

$$\max_{t \in [0,1]} ||\beta'(v(t)) - \alpha'(u(t))|| - d(\alpha, \beta) < \epsilon.$$

Define $u_\epsilon, v_\epsilon : [0,1] \to [0,1]$ by

$$u_\epsilon(t) = \frac{u(t) + \epsilon t}{1+\epsilon} \quad \text{and} \quad v_\epsilon(t) = \frac{v(t) + \epsilon t}{1+\epsilon}.$$

Now $u_\epsilon, v_\epsilon \in F$ because they both are a sum of a nondecreasing surjection $[0,1] \to [0, 1/(1+\epsilon)]$ and an increasing function $[0,1] \to [0, \epsilon/(1+\epsilon)]$. Furthermore, for all $t \in [0,1]$,

$$\begin{aligned} |u_\epsilon(t) - u(t)| &= \left| \frac{u(t) + \epsilon t}{1+\epsilon} - u(t) \right| \\ &= \left| \frac{u(t) + \epsilon t - (1+\epsilon)u(t)}{1+\epsilon} \right| \\ &= \epsilon \left| \frac{t - u(t)}{1+\epsilon} \right| \leq \epsilon. \end{aligned}$$

Similarly $|v_\epsilon(t) - v(t)| < \epsilon$ for all $t \in [0,1]$. Now if we choose $f = v_\epsilon \circ u_\epsilon^{-1} \in F$, by reparameterizing the maximum we get

$$\begin{aligned} \max_{t\in[0,1]} ||\beta'(f(t)) - \alpha'(t)|| &= \max_{t\in[0,1]} ||\beta'(v_\epsilon(u_\epsilon^{-1}(t))) - \alpha'(u_\epsilon(u_\epsilon^{-1}(t)))|| \\ &= \max_{t\in[0,1]} ||\beta'(v_\epsilon(t)) - \alpha'(u_\epsilon(t))|| \end{aligned}$$

Because α and β are 1-Lipschitz, α' and β' are $|\alpha|$- and $|\beta|$-Lipschitz, respectively. Therefore by using the triangle inequality, we get that

$$\left| \max_{t\in[0,1]} ||\beta'(v_\epsilon(t)) - \alpha'(u_\epsilon(t))|| - \max_{t\in[0,1]} ||\beta'(v(t)) - \alpha'(u(t))|| \right| \leq (|\alpha| + |\beta|)\epsilon.$$

By putting this together with the defining condition of u and v and using the triangle inequality, we get that

$$\left| \max_{t\in[0,1]} ||\beta'(v_\epsilon(t)) - \alpha'(u_\epsilon(t))|| - d(\alpha,\beta) \right| \leq (|\alpha| + |\beta| + 1)\epsilon.$$

Therefore

$$\begin{aligned} d'(\alpha,\beta) &\leq \max_{t\in[0,1]} ||\beta'(f(t)) - \alpha'(t)|| \\ &= \max_{t\in[0,1]} ||\beta'(v_\epsilon(t)) - \alpha'(u_\epsilon(t))|| \\ &\leq d(\alpha,\beta) + (|\alpha| + |\beta| + 1)\epsilon, \end{aligned}$$

and as we take $\epsilon \to 0$, we prove that $d' \leq d$. □

Remark. In both definitions of d we use the unit interval parameterizations of the paths for simplicity of notation. However, we can also write the definition of $d(\alpha,\beta)$ for $\alpha, \beta \in \mathcal{P}$ simply as

$$d(\alpha,\beta) = \inf_{\substack{u:[0,1]\to[0,|\alpha|] \text{ surjective} \\ v:[0,1]\to[0,|\beta|] \text{ surjective}}} \max_{t\in[0,1]} ||\beta(v(t)) - \alpha(u(t))||.$$

Also if $|\alpha| > 0$ or $|\alpha| = |\beta| = 0$, we can rewrite the alternative definition as

$$d(\alpha, \beta) = \inf_{f:[0,|\alpha|]\to[0,|\beta|] \text{ bijective}} \max_{t\in[0,|\alpha|]} ||\beta(f(t)) - \alpha(t)||.$$

Now that we have both definitions in use, we are ready to prove that this distance d is a metric.

Theorem 3.3. *The distance d defined by Definition 3.1 is a metric in $\mathcal{P}$.*

Proof. ***Symmetry***. We get that $d(\alpha, \beta) = d(\beta, \alpha)$ for all $\alpha, \beta \in \mathcal{P}$ directly from the symmetry of the definition, by just swapping u and v.
Subadditivity. Let $\alpha, \beta, \gamma \in \mathcal{P}$. We use the alternative definition of Lemma 3.2. Let $f, g \in F$. By reparameterizing (which is allowed, because $f[0,1] = [0,1]$), combining maximums and using the triangle inequality, we get that

$$\begin{aligned}
&\max_{t\in[0,1]} ||\beta'(f(t)) - \alpha'(t)|| + \max_{t\in[0,1]} ||\gamma'(g(t)) - \beta'(t)|| \\
&= \max_{t\in[0,1]} ||\beta'(f(t)) - \alpha'(t)|| + \max_{t\in[0,1]} ||\gamma'(g(f(t))) - \beta'(f(t))|| \\
&\geq \max_{t\in[0,1]} \left(||\beta'(f(t)) - \alpha'(t)|| + ||\gamma'(g(f(t))) - \beta'(f(t))||\right) \\
&\geq \max_{t\in[0,1]} ||\gamma'(g(f(t))) - \alpha'(t)||.
\end{aligned}$$

Because $g \circ f \in F$, this is greater or equal to $d(\alpha, \gamma)$. Because this holds for all $f, g \in F$, it holds also for the infimum over all such f and g, so the triangle inequality $d(\alpha, \beta) + d(\beta, \gamma) \geq d(\alpha, \gamma)$ holds.
Identity. By setting $\beta := \alpha$ and u and v to identity mappings in the definition, all differences cancel out to zero, so $d(\alpha, \alpha) = 0$ for all $\alpha \in \mathcal{P}$.
Separation. Let $\alpha, \beta \in \mathcal{P}$ such that $d(\alpha, \beta) = 0$. We want to prove that $\alpha = \beta$. Without loss of generality we may assume that $|\alpha| \geq |\beta|$. If $|\beta| = 0$, then β is a constant path, and clearly now α has to also be a constant path at the same point, so $\alpha = \beta$. Assume that $|\beta| > 0$. Let $\epsilon > 0$ and $t \in [0, |\alpha|]$. Choose $S \subset [0, t]$ and $S' \subset [t, |\alpha|]$ such that

$$|\alpha|_{[0,t]} - \sum_{k=1}^{|S|-1} ||\alpha(S_{[k+1]}) - \alpha(S_{[k]})|| < \epsilon, \tag{1}$$

$$|\alpha|_{[t,|\alpha|]} - \sum_{k=1}^{|S'|-1} ||\alpha(S'_{[k+1]}) - \alpha(S'_{[k]})|| < \epsilon. \tag{2}$$

By the alternative definition of d in Lemma 3.2, we can choose $f \in F$ such that

$$\max_{t\in[0,1]} ||\beta'(f(t)) - \alpha'(t)|| < \epsilon/n,$$

where $n = |S| + |S'|$. Now if we define function $g : [0, |\alpha|] \to [0, |\beta|]$ by $g(t) = |\beta| f(t/|\alpha|)$, we get that g is an increasing bijection such that

$$\max_{t \in [0,|\alpha|]} ||\beta(g(t)) - \alpha(t)|| < \epsilon/n.$$

Because g is increasing, we get lower bounds for the lengths $|\beta|_{[0,g(t)]}$ and $|\beta|_{[g(t),|\beta|]}$:

$$\begin{aligned} |\beta|_{[0,g(t)]} &\geq \sum_{k=1}^{|S|-1} ||\beta(g(S_{[k+1]})) - \beta(g(S_{[k]}))||, \\ |\beta|_{[g(t),|\beta|]} &\geq \sum_{k=1}^{|S'|-1} ||\beta(g(S'_{[k+1]})) - \beta(g(S'_{[k]}))||. \end{aligned}$$

Because $||\beta(g(t)) - \alpha(t)|| < \epsilon/n$, by approximating $\beta(g(t))$ by $\alpha(t)$ and using the triangle inequality we get weaker lower bounds using values of α:

$$\begin{aligned} |\beta|_{[0,g(t)]} + 2\epsilon &\geq \sum_{k=1}^{|S|-1} ||\alpha(S_{[k+1]}) - \alpha(S_{[k]})||, \\ |\beta|_{[g(t),|\beta|]} + 2\epsilon &\geq \sum_{k=1}^{|S'|-1} ||\alpha(S'_{[k+1]}) - \alpha(S'_{[k]})||. \end{aligned}$$

Combining these with (1) and (2), we get that $|\alpha|_{[0,t]} - |\beta|_{[0,g(t)]} < 3\epsilon$ and $|\alpha|_{[t,|\alpha|]} - |\beta|_{[g(t),|\beta|]} < 3\epsilon$. Because α and β are paths, these translate to

$$\begin{aligned} t - g(t) &< 3\epsilon, \\ |\alpha| - t - |\beta| + g(t) &< 3\epsilon. \end{aligned}$$

If we choose $t = 0$, then for all $\epsilon > 0$ we get that $|\alpha| - |\beta| < 3\epsilon$, and because $|\alpha| \geq |\beta|$, we get that $|\alpha| = |\beta|$. Now the inequalities yield that $|g(t) - t| < 3\epsilon$. We can bound $||\beta(t) - \alpha(t)||$ using the triangle inequality:

$$\begin{aligned} ||\beta(t) - \alpha(t)|| &= ||\beta(t) - \beta(g(t)) + \beta(g(t)) - \alpha(t)|| \\ &\leq ||\beta(t) - \beta(g(t))|| + ||\beta(g(t)) - \alpha(t)|| \\ &\leq |g(t) - t| + \epsilon/n \\ &\leq 6\epsilon + \epsilon = 7\epsilon, \end{aligned}$$

where the middle inequality follows from the fact that β is 1-Lipschitz, as proved in Theorem 2.12. Now we have proved that for all $\epsilon > 0$ and $t \in [0, |\alpha|]$, $||\beta(t) - \alpha(t)|| < 7\epsilon$, and therefore $\alpha = \beta$. □

This distance metric d we defined has the natural property that distance between two paths can be overestimated by splitting them to pieces and taking the larger of the distances of the corresponding pieces.

Theorem 3.4. *Let paths $\alpha, \alpha_1, \alpha_2, \beta, \beta_1, \beta_2 \in \mathcal{P}$ be such that $\alpha = \alpha_1\alpha_2$ and $\beta = \beta_1\beta_2$. Now*

$$d(\alpha, \beta) \leq \max\{d(\alpha_1, \beta_1), d(\alpha_2, \beta_2)\}.$$

Proof. Let $\epsilon > 0$. Choose u_k, v_k for $k \in \{1, 2\}$ such that

$$\max_{t\in[0,1]} ||\beta_k'(v_k(t)) - \alpha_k'(u_k(t))|| - d(\alpha_k, \beta_k) < \epsilon. \tag{3}$$

Now define $u : [0, 1] \to [0, |\alpha|]$ and $v : [0, 1] \to [0, |\beta|]$ by

$$u(t) = \begin{cases} |\alpha_1|u_1(2t), & \text{if } t \leq 1/2 \\ |\alpha_1| + |\alpha_2|u_2(2t-1), & \text{if } t > 1/2 \end{cases}$$

$$v(t) = \begin{cases} |\beta_1|v_1(2t), & \text{if } t \leq 1/2 \\ |\beta_1| + |\beta_2|v_2(2t-1), & \text{if } t > 1/2. \end{cases}$$

Clearly now in both definitions the branches match at $t = 1/2$ and both branches are continuous nondecreasing functions. Also $u(0) = v(0) = 0$, $u(1) = |\alpha|$ and $v(1) = |\beta|$. Therefore u and v are surjections. Now we get a upper bound for the distance of α and β:

$$\begin{aligned} d(\alpha, \beta) &\leq \max_{t\in[0,1]} ||\beta(v(t)) - \alpha(u(t))|| \\ &= \max\left\{\max_{t\in[0,\frac{1}{2}]} ||\beta(v(t)) - \alpha(u(t))||, \max_{t\in[\frac{1}{2},1]} ||\beta(v(t)) - \alpha(u(t))||\right\} \end{aligned}$$

Denote the first argument to max by A and the second by B. By expanding them, we get

$$\begin{aligned} A &= \max_{t\in[0,\frac{1}{2}]} ||\beta(|\beta_1|v_1(2t)) - \alpha(|\alpha_1|u_1(2t))|| \\ &= \max_{t\in[0,1]} ||\beta_1'(v_1(t)) - \alpha_1'(u_1(2t))||. \\ B &= \max_{t\in[\frac{1}{2},1]} ||\beta(|\beta_1| + |\beta_2|v_2(2t-1)) - \alpha(|\alpha_1| + |\alpha_2|u_2(2t-1))|| \\ &= \max_{t\in[0,1]} ||\beta_2'(v_2(t)) - \alpha_2'(u_2(t))||. \end{aligned}$$

From (3) we get that $|d(\alpha_1, \beta_1) - A| < \epsilon$ and $|d(\alpha_2, \beta_2) - B| < \epsilon$, which yields the inequality

$$d(\alpha, \beta) \leq \max\{d(\alpha_1, \beta_1), d(\alpha_2, \beta_2)\} + \epsilon.$$

Because this holds for all $\epsilon > 0$, the proof is complete. □

3.2 Basic properties

Now that we have a metric d in the space of paths, we can proceed to define locally shortest paths:

Definition 3.5. Let $a, b \in P$ and $\gamma \in \mathcal{P}_{ab}$. γ is a *locally shortest path*, if it is a local minimum of the path distance function $\alpha \mapsto |\alpha|$ in the set $\mathcal{P}_{ab}$, i.e. there exists some $\epsilon > 0$ such that for all $\alpha \in \mathcal{P}_{ab}$ with $d(\gamma, \alpha) < \epsilon$ it holds that $|\alpha| \geq |\gamma|$. We denote the set of *locally shortest paths* from a to b by $\mathcal{L}_{ab}$, and the set of all locally shortest paths $\bigcup_{a,b \in P} \mathcal{L}_{ab}$ by $\mathcal{L}$.

In the design of shortest path algorithms, the fact that all subpaths of shortest paths are also shortest paths is often crucial. It turns out that locally shortest paths have the same property.

Theorem 3.6. *Let $\gamma \in \mathcal{L}_{xy}$ and α be a subpath of γ. Then $\alpha \in \mathcal{L}$.*

Proof. Let $a, b \in P$ such that $\alpha \in \mathcal{P}_{ab}$. Because α is a subpath of γ, there exists $\xi \in \mathcal{P}_{xa}$ and $\phi \in \mathcal{P}_{by}$ such that $\gamma = \xi\alpha\phi$. Assume the contraposition, i.e. $\alpha \notin \mathcal{L}$. Then for all $\epsilon > 0$ there exists $\alpha_\epsilon \in \mathcal{P}_{ab}$ such that $d(\alpha, \alpha_\epsilon) < \epsilon$ and $|\alpha_\epsilon| < |\alpha|$. Define $\gamma_\epsilon = \xi\alpha_\epsilon\phi$. Now

$$|\gamma_\epsilon| = |\xi| + |\alpha_\epsilon| + |\phi| < |\xi| + |\alpha| + |\phi| = |\gamma|,$$

and by Theorem 3.4,

$$\begin{aligned} d(\gamma_\epsilon, \gamma) &= d(\xi\alpha_\epsilon\phi, \xi\alpha\phi) \\ &= d(\xi\alpha_\epsilon, \xi\alpha) + d(\phi, \phi) \\ &= d(\xi, \xi) + d(\alpha_\epsilon, \alpha) + d(\phi, \phi) \\ &< \epsilon. \end{aligned}$$

Thus $\gamma \notin \mathcal{L}_{xy}$, which is a contradiction. $\square$

The definition of locally shortest paths means that a locally shortest path is a path in the domain that cannot be made shorter by using small perturbations without moving the endpoints. Physically, this means that the path, thought of as a piece of string, is pulled taut in the domain (see Figure 3.1). As the domain P is a polygonal domain, intuitively it would seem that a string pulled taut in P always takes the form of a polygonal chain, turning only in the vertices of the domain. We will prove that this holds. We will also prove the inverse result stating that such paths are always locally shortest, provided that the paths always turn towards the obstacle (or *make reflex turns*, as we will define in Definition 3.8). Note that the theory we have established so far works for general domains P. However, that is not the case with the next theorem: for the local considerations in its proof we need the following properties that follow from the fact that P is a polygonal domain.

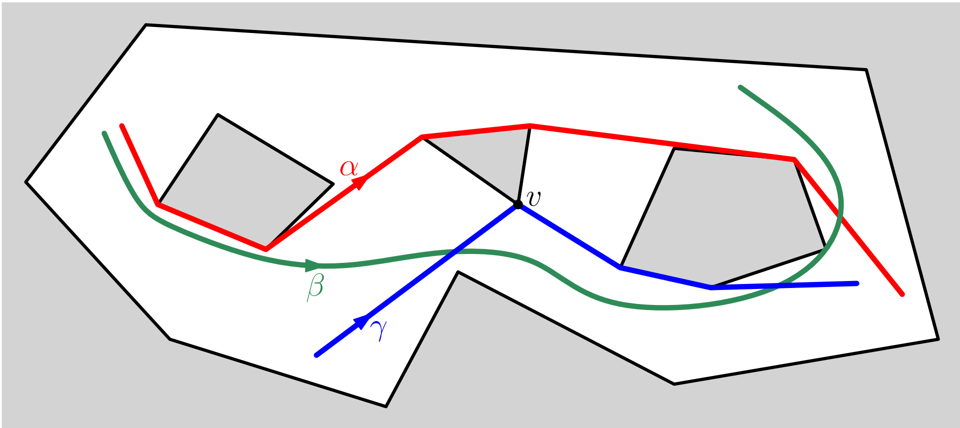


Figure 3.1: Among the three paths in this figure, only α is a locally shortest path. This can be seen using the intuitive definition for locally shortest paths, which is that a path is locally shortest if it is *pulled taut*. More exactly, if the path is thought of as a piece of string in the domain P, and it is pulled from both ends, it should not change. As α always turns only at obstacle polygon vertices, and always towards the obstacle, there is nowhere it can move when it is pulled from the ends. γ is almost like that, but it turns away from the obstacle at vertex v, which means pulling it tighter will move it away from vertex v.

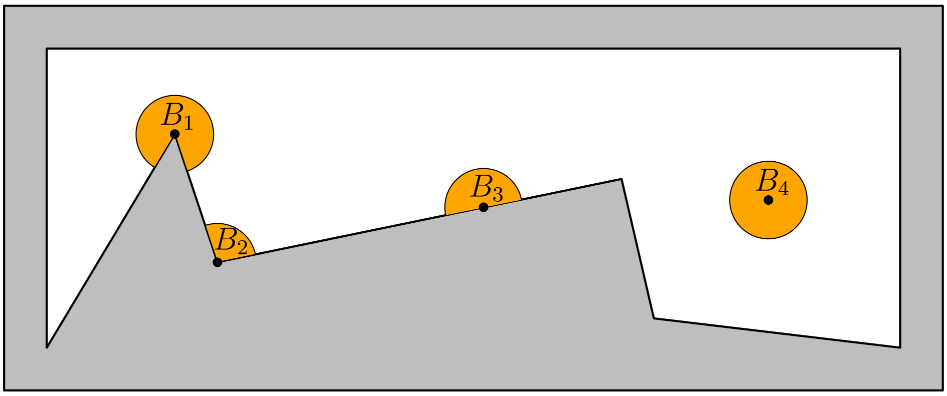


Figure 3.2: B_1 and B_2 are neighborhoods of polygon vertices, B_3 is a neighborhood of an edge point of P, and B_4 is a neighborhood of an interior point of P. B_1 and B_2 are disk sectors, B_3 is a half-disk and B_4 is a disk.

Lemma 3.7. *Let $x \in P$. Now there exists $r > 0$, such that for all $0 < \epsilon < r$, the ϵ-neighborhood of x in P, that is,*

$$B_P(x, \epsilon) = B(x, \epsilon) \cap P = \{y \in P \mid ||x - y|| < \epsilon\},$$

is as follows:

- *If x is a vertex ($x \in V$), then $B_P(x, \epsilon)$ is a disk sector (for example B_1 and B_2 in Figure 3.2).*
- *If x is on an edge of the domain ($x \in \partial P \setminus V$), then $B_P(x, \epsilon)$ is a half-disk (for example B_3 in Figure 3.2).*
- *If x is in the interior of the domain ($x \notin \partial P$), then $B_P(x, \epsilon)$ is the disk $B_{\mathbb{R}^2}(x, r)$ (for example B_4 in Figure 3.2).*

In any case, $B_P(x, r)$ is a disk or a disk sector.

Proof. If $x \in V$, we can choose r as the distance to the closest point on the set of edges of ∂P not incident to x. As that set is compact, the minimum exists and is positive.

Similarly, if $x \in \partial P \setminus V$, we can choose r as the distance to the closest point on the set of edges with the edge x lies on removed.

If x is in the interior of the domain, the existence of such $r > 0$ follows directly from the definition of open sets in metric spaces. □

We also need terminology on how a polygonal chain goes through a vertex.

Definition 3.8. Let γ be a polygonal chain in P and $t \in (0, |\gamma|)$ such that $\gamma(t) = v \in V$. We denote the angle between the incoming and outgoing path γ at t that is inside P by $\angle(\gamma, t)$ (see Figure 3.3). If $\angle(\gamma, t) \geq 180°$, we say that γ makes a *reflex turn* at t. If the choice of value of t among $\gamma^{-1}\{v\}$ is clear from the context, we also say that γ makes a reflex turn at v, and denote $\angle(\gamma, v) = \angle(\gamma, t)$.

Theorem 3.9. *If $\gamma \in \mathcal{P}_{ab}$, then $\gamma \in \mathcal{L}_{ab}$ if and only if $\gamma = [a, v_1, v_2, \ldots, v_n, b]$ with some vertices $v_1, v_2, \ldots, v_n \in V$ such that for all $t \in (0, |\gamma|)$ for which $\gamma(t) \in V$, γ makes a reflex turn at t ($\angle(\gamma, t) \geq 180°$).*

Proof. ***Direction*** $\Rightarrow$. Assume that $\gamma \in \mathcal{L}_{ab}$. First we prove that γ is a polygonal chain, and after that we prove that all turns are reflex turns at polygon vertices. Let X be the set of $t \in [0, |\gamma|]$ such that $\gamma_{[0,t]}$ is a polygonal chain. Let $t = \sup X$ and $x = \gamma(t)$. To prove that γ is a polygonal chain, we only need to prove that $t = |\gamma|$. Assume the contrary: $t < |\gamma|$.

By Lemma 3.7, there exists $r > 0$ such that for all $0 < \epsilon < r$, $B_P(x, \epsilon)$ is a disk sector or a disk centered at x. Denote $t' = \max\{t - \epsilon/2, 0\}$ and

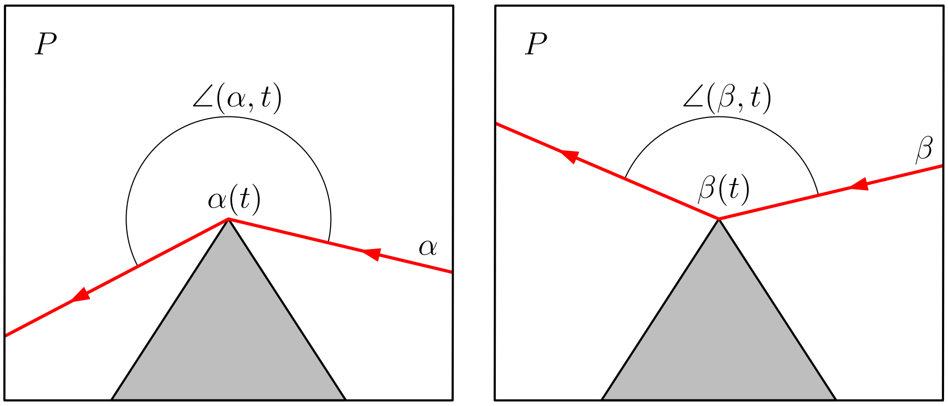
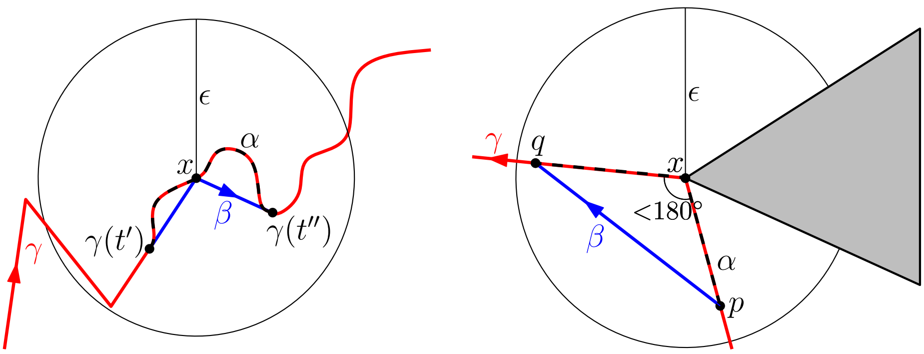


Figure 3.3: $\angle(\gamma, t)$ is the angle between the incoming and the outgoing path γ at t that is inside P. Now $\angle(\alpha, t) > 180°$ and $\angle(\beta, t) < 180°$, which means that α makes a reflex turn at t, but β does not.



(a) For sufficiently small $\epsilon > 0$, the ϵ-neighborhood of $x = \gamma(t)$ is a disk or a half-disk (disk in the figure). Now α defined as a subpath of γ around t is not a polygonal chain, so we can shortcut it using a polygonal chain β, proving the contradiction $\gamma \notin \mathcal{L}$.

(b) Let $x = \gamma(t)$ be a turn vertex of γ such that either $x \notin V$ or γ does not make reflex turn at t (in the figure). In both cases, we may take a subpath α of γ in an arbitrarily small $B_P(x, \epsilon)$ and shortcut it using a polygonal chain β proving the contradiction $\gamma \notin \mathcal{L}$.

Figure 3.4: Visual demonstration of the proof in direction $\Rightarrow$ of Theorem 3.9.

$t'' = \min\{t + \epsilon/2, |\gamma|\}$. Because γ is 1-Lipschitz by Theorem 2.12, we get that

$$\gamma[t', t''] \subset B_P(x, \epsilon).$$

Define $\alpha = \gamma_{[t',t'']}$ and $\beta = [\gamma(t'), x, \gamma(t'')]$ (see Figure 3.4a). β is contained in $B_P(x, \epsilon)$ because all the points in it are visible from x. Because α and β are contained in a disk of radius ϵ, any two points in α and β have distance at most 2ϵ. By the definition of t, $\gamma_{[0,t']}$ is a polygonal chain and $\gamma_{[0,t'']}$ is not, and therefore α is not a polygonal chain. Because β is a polygonal chain, $\alpha \neq \beta$ and thus by Theorem 2.12, $|\beta| < |\alpha| = t'' - t' \leq \epsilon$. Therefore $d(\alpha, \beta) < 3\epsilon$. Using Theorem 3.4, we can compute

$$d(\gamma, \gamma_{[0,p]}\beta\gamma_{[q,|\gamma|]}) = d(\alpha, \beta) < 3\epsilon.$$

Thus for all $\epsilon > 0$ we find a shorter path than γ in $\mathcal{P}_{ab}$ with distance less than 3ϵ, so $\gamma \notin \mathcal{L}$, which is a contradiction. Thus $t = |\gamma|$ and γ is a polygonal chain.

Now that we have proved that γ is a polygonal chain, it remains to prove that it only turns at polygon vertices, and that the turns are reflex turns. Let $\gamma = [\gamma(t_0) = a, \gamma(t_1), \ldots, \gamma(t_n) = b]$ be the shortest representation of γ as a polygonal chain. Assume that for some $k \in \{1, 2, \ldots, n-1\}$, at least one of the following claims does not hold: $\gamma(t_k) \in V$ and γ makes a reflex turn at t_k. Denote $x = \gamma(t_k)$.

By Lemma 3.7, there exists $r > 0$ such that for all $0 < \epsilon < r$, $B_P(x, \epsilon)$ is a disk sector. Let $\epsilon = \min\{r/2, t_{k+1} - t_k, t_k - t_{k-1}\}$. Denote $t' = t_k - \epsilon/2$ and $t'' = t_k + \epsilon/2$. Define $\alpha = \gamma_{[t',t'']}$. Now $\alpha = [p, x, q]$, where p is inside the segment $[\gamma(t_{k-1}), \gamma(t_k)]$ and q is inside the segment $[\gamma(t_k), \gamma(t_{k+1})]$. Because α is 1-Lipschitz, it is contained in $B_P(x, \epsilon)$. Define $\beta = [p, q]$. If $x \notin V$, $B_P(x, \epsilon)$ is either a disk or a half-disk and therefore convex, so β is contained in $B_P(x, \epsilon)$. Otherwise, $x \in V$ and $\angle(\gamma, t_k) < 180°$. In this case too, β is contained in $B_P(x, \epsilon)$ (see Figure 3.4b). Because both α and β are contained in $B_P(x, \epsilon)$, and have length at most 2ϵ, they have distance at most 4ϵ. Using Theorem 3.4, we can compute

$$d(\gamma, \gamma_{[0,p]}\beta\gamma_{[q,|\gamma|]}) = d(\alpha, \beta) < 4\epsilon.$$

Furthermore, $|\beta| < |\alpha|$, because by the minimality assumption points $\gamma(t_{k-1}), \gamma(t_k), \gamma(t_{k+1})$ are not collinear, and thus points p, x, q are not collinear either. We have found for all $\epsilon > 0$ a shorter path that is closer than 4ϵ to γ, which yields a contradiction $\gamma \notin \mathcal{L}$.

Direction $\Leftarrow$. Let $\gamma = [a, v_1, v_2, \ldots, v_n, b]$ with $v_1, v_2, \ldots, v_n \in V$ such that for all $t \in (0, |\gamma|)$ for which $\gamma(t) \in V$, γ makes a reflex turn at t. We can assume that the list of vertices does not contain repetitions, i.e. $a \neq v_1 \neq v_2 \neq \ldots \neq v_n \neq b$. Let $0 < t_1 < t_2 < \ldots < t_n < |\gamma|$ be the parameters corresponding to the list of vertices, i.e. $\gamma(t_k) = v_k$ for all

$k \in \{0, 1, \ldots, n+1\}$. For convenience, denote $v_0 = a$ and $v_{n+1} = b$. Now $\angle(\gamma, t_k) \geq 180°$ for all $k \in \{1, \ldots, n\}$. We may assume that $\angle(\gamma, t_k) > 180°$ for all $k \in \{1, \ldots, n\}$, because otherwise we can remove v_k from the list of vertices: $\gamma = [a, v_1, \ldots, v_{k-1}, v_{k+1}, \ldots, v_n, b]$.

Let H_k be the bisector halfline drawn for the reflex angle that γ makes at t_k for all $k \in \{1, \ldots, n\}$, i.e. larger of the angles $\angle v_{k-1} v_k v_{k+1}$ and $\angle v_{k+1} v_k v_{k-1}$. For convenience, define $H_0 = \{a\}$ and $H_{n+1} = \{b\}$.

We will prove that there exists $\epsilon > 0$ such that for all $\alpha \in \mathcal{P}_{ab}$ with $d(\alpha, \gamma) < \epsilon$, $|\alpha| \geq |\gamma|$. First we prove that for sufficiently small $\epsilon > 0$, α goes through the halflines $H_1, H_2, \ldots, H_n$ in that order, i.e. there exists parameters $0 = x_0 < x_1 < \ldots < x_{n+1} = |\alpha|$ such that $\alpha(x_k) \in H_k$ for all $k \in \{0, \ldots, n+1\}$ (see Figure 3.5a).

Choose radius $r > 0$ such all P-disks $B_P(v, r)$ for $v \in \{v_1, \ldots, v_n\}$ are disjoint disk sectors that do not contain a or b. The existence of such r follows from Lemma 3.7 and the fact that $\{a, v_1, \ldots, v_n, b\}$ is a discrete set. Similarly, we can choose $\epsilon > 0$ such that for all $k \in \{1, 2, \ldots, n\}$, $B_P(\gamma(t_k - r/2), \epsilon)$ and $B_P(\gamma(t_k + r/2), \epsilon)$ are disjoint disks/half-disks that do not contain v_k (see Figure 3.5b). If $\alpha \in \mathcal{P}_{ab}$ and $d(\alpha, \gamma) < \epsilon$, then by the alternative definition of d in Lemma 3.2, there exists an increasing bijection $f : [0, |\gamma|] \to [0, |\alpha|]$ such that

$$\max_{t \in [0,1]} ||\alpha(f(t)) - \gamma(t)|| < \epsilon.$$

Denote $I_k = \left[t_k - \frac{r}{2}, t_k + \frac{r}{2}\right]$. Now

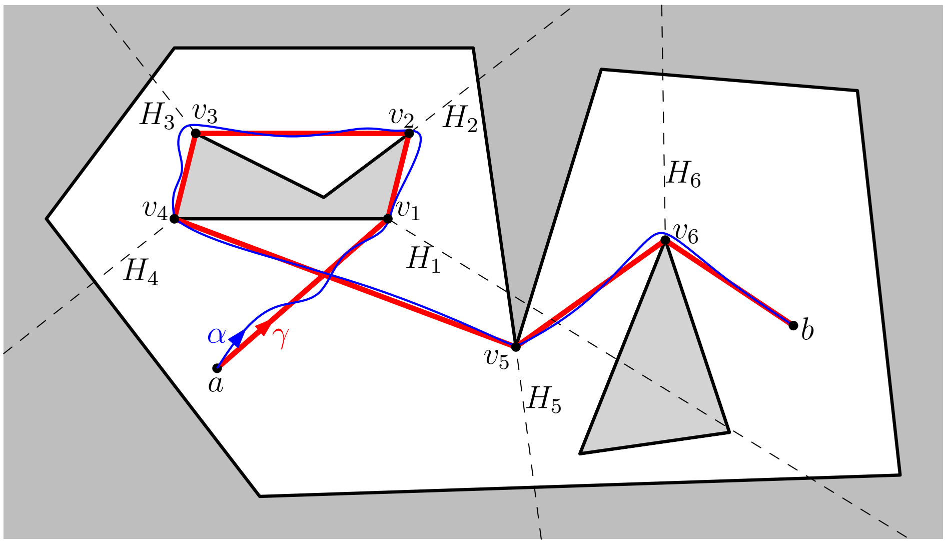
$$0 < t_1 - \frac{r}{2} < t_1 + \frac{r}{2} < \ldots < t_n - \frac{r}{2} < t_n + \frac{r}{2} < |\gamma|,$$

and thus it is enough to prove that for all $k \in \{1, \ldots, n\}$ there exists $t \in I_k$ such that $\alpha(f(t)) \in H_k$, because then we know that α visits bisectors $H_1, H_2, \ldots, H_n$ in that order. Because γ is 1-Lipschitz, for all $t \in I_k$ it holds that $||\gamma(t) - v_k|| < r/2$, and therefore by the triangle inequality, $||\alpha(f(t)) - v_k|| < r/2 + \epsilon \leq r$. Thus $\alpha \circ f(I_k) \subset B_P(v_k, r)$. Furthermore, $\alpha(f(t_k - \frac{r}{2})) \in B_P(\gamma(t_k - \frac{r}{2}), \epsilon)$ and $\alpha(f(t_k + \frac{r}{2})) \in B_P(\gamma(t_k + \frac{r}{2}), \epsilon)$, and those disks are separated by the bisector halfline H_k in $B_P(v_k, r)$, which yields that there exists $t \in I_k$ such that $\alpha(f(t)) \in H_k$ (see Figure 3.5b for more details).

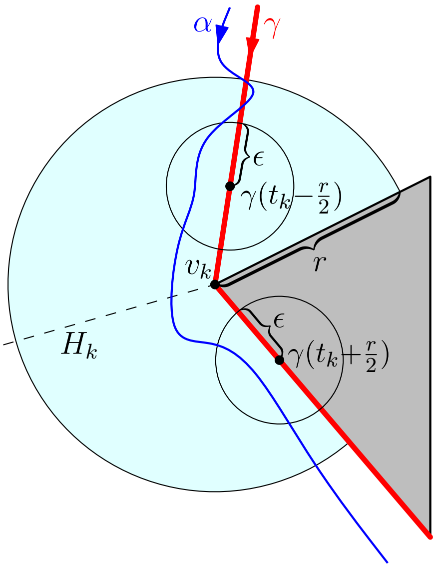
Now we get the result $|\alpha| \geq |\gamma|$ for all $\alpha \in \mathcal{L}_{ab}$ such that $d(\alpha, \gamma) < \epsilon$, because if $k \in \{0, \ldots, n\}$, $x \in H_k$ and $y \in H_{k+1}$, then $||y - x|| \geq ||v_{k+1} - v_k||$ (see Figure 3.5c for more details), which by the 1-Lipschitz property of α means that

$$|\alpha| \geq \sum_{k=0}^{n} ||\alpha(x_{k+1}) - \alpha(x_k)|| \geq \sum_{k=0}^{n} ||v_{k+1} - v_k|| = |\gamma|.$$

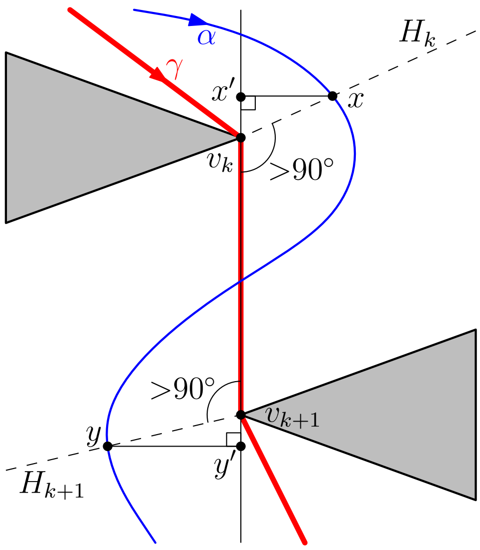
□



(a) If a path $\gamma = [a, v_1, v_2, \ldots, v_n, b]$ makes only reflex turns, if we draw a bisector halfline H_k for the angle γ makes at every v_k, we notice that any path $\alpha \in \mathcal{P}_{ab}$ that is sufficiently close to γ in d has to traverse the bisectors in order $H_1, H_2, \ldots, H_n$.



(b) There exists radii $r, \epsilon > 0$ such that the neighborhood of any turn vertex $v_k = \gamma(t_k)$ looks like the figure. H_k splits $B_P(v_k, r)$ into two components, both containing exactly one of the disks $B_P(\gamma(t_k - r/2), \epsilon)$ and $B_P(\gamma(t_k + r/2), \epsilon)$. If $d(\alpha, \gamma) < \epsilon$, then α must pass through both disks, and thus it must cross H_k at $\alpha(t)$ for some $t \in [t_k - r/2, t_k + r/2]$.



(c) Let $x \in H_k$ and $y \in H_{k+1}$ be the points where α hits the bisectors. Now because γ makes only reflex turns, both angles $\angle v_{k+1} v_k x$ and $\angle v_k v_{k+1} y$ are obtuse, and therefore the projections x' and y' of the points x and y to the line $v_k v_{k+1}$ are external. Because projecting points to a line can only decrease their distances, $||x - y|| \geq ||x' - y'|| \geq ||v_{k+1} - v_k||$.

Figure 3.5: Visual demonstration of the proof of Theorem 3.9 in direction $\Leftarrow$.

As a corollary of this theorem we also get that the problem of listing the paths of $\mathcal{L}_{ab}$ ordered by length is well-posed.

Theorem 3.10. *Let $a, b \in P$. Then the set $S_R = \{\gamma \in \mathcal{L}_{ab} \mid |\gamma| \leq R\}$ is finite for all $R \in \mathbb{R}$.*

Proof. Define $L = \min\{||y-x|| \mid x, y \in V, x \neq y\}$. Because V is finite, $L > 0$. Let $\gamma \in \mathcal{L}_{ab}$ and $|\gamma| \leq R$. Now by Theorem 3.9, γ can be written as

$$\gamma = [a, v_1, v_2, \ldots, v_n, b]$$

for some $v_1, \ldots, v_n \in V$. We may assume that the chain is written without repetitions ($v_1 \neq v_2 \neq \ldots \neq v_n$), because they can be removed without changing the path. Now

$$\begin{aligned} |\gamma| &= |[a, v_1]| + |[v_1, v_2]| + |[v_2, v_3]| + \ldots + |[v_n, b]| \\ &\geq |[v_1, v_2]| + |[v_2, v_3]| + \ldots + |[v_{n-1}, v_n]| \\ &\geq (n-1)L. \end{aligned}$$

We know that $|\gamma| \leq R$, so $n \leq R/L + 1$. Therefore

$$S_R \subset \{[a, v_1, v_2, \ldots, v_n, b] \in \mathcal{P}_{ab} \mid n \leq R/L + 1 \text{ and } v_1, v_2, \ldots, v_n \in V\}.$$

Therefore S_R has size at most $|V|^0 + |V|^1 + \ldots + |V|^{\lfloor R/L+1 \rfloor}$, which is finite. $\square$

3.3 $\mathcal{L}_s$ as a covering space

Define the set $\mathcal{L}_s$ of locally shortest paths from a fixed source point s by

$$\mathcal{L}_s = \bigcup_{x \in P} \mathcal{L}_{sx}.$$

In this subsection, we will consider the structure of the set $\mathcal{L}_s$. We will prove that for all locally shortest paths $\gamma \in \mathcal{L}_{sx}$, by moving the endpoint x of γ continuously, the path γ can always be changed continuously with respect to distance d while keeping it in $\mathcal{L}_s$, and locally this new path is unique for the choice of endpoint. In the language of topology, this means that the function that maps a path to its endpoint $\gamma \mapsto \gamma(|\gamma|)$ is a *covering map* from $\mathcal{L}_s$ to P, and thus $\mathcal{L}_s$ is a *covering space* of P. To give an intuitive understanding of this covering space, Figure 3.6 contains a depiction of the structure of the set $\mathcal{L}_s$ as a two-dimensional surface in a three-dimensional space.

Definition 3.11. Let X and Y be metric spaces. Then a mapping $p : X \to Y$ is a *covering map*, if it is surjective and for every $y \in Y$ there exists $r_y > 0$ such that $p^{-1}B_Y(y, r_y)$ can be written as $\bigcup \mathcal{U}_y$, where $\mathcal{U}_y$ is a collection of pairwise disjoint open sets, and for all $U \in \mathcal{U}_y$, $p|_U$ is a homeomorphism $U \to B_Y(y, r_y)$. In this case, X is called the *covering space* (of Y) and Y is called the *base space*.

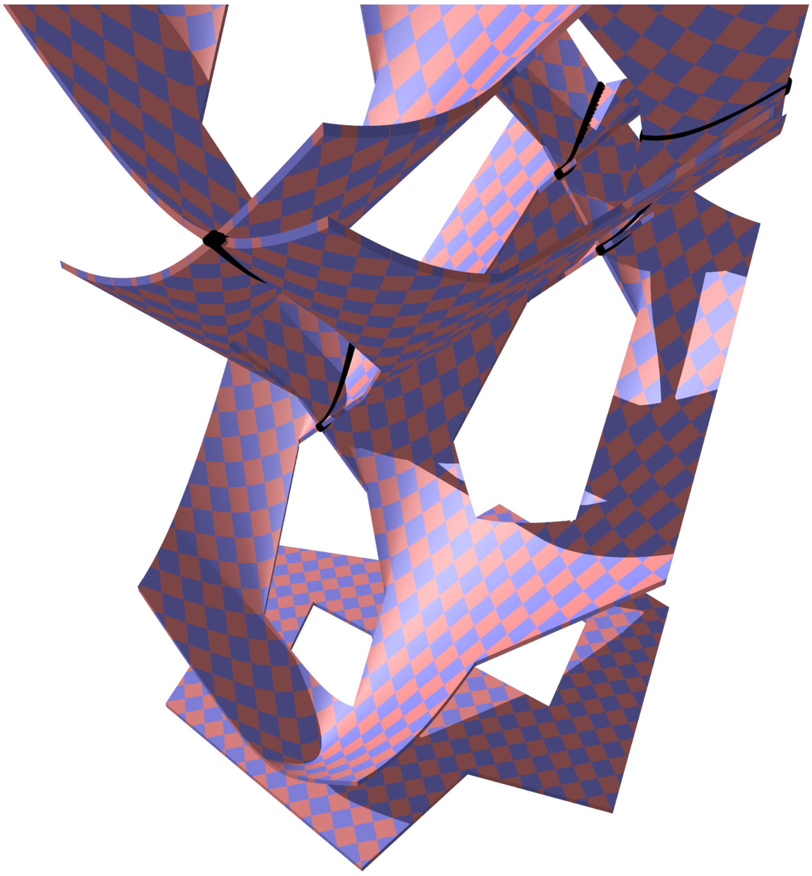


Figure 3.6: A three-dimensional projection of the set of locally shortest paths from a fixed point s, where each $\gamma \in \mathcal{L}_s$ with endpoint $e(\gamma) = (x, y)$ is mapped to the point $(x, y, |\gamma|)$. In this example, we use a simple polygonal domain with two holes as the domain P. The domain P is drawn to the base plane $z = 0$. The set of points forms an infinite surface with 45° inclination, constrained by the domain P, and covering it many times from above. The surface crosses itself in points marked in black. In these points, there are multiple locally shortest paths that map to the same point, i.e. have the same endpoint and the same length. In Theorem 3.13, we prove that $\mathcal{L}_s$ is a covering space of P, which means that locally the set $\mathcal{L}_s$ looks like the domain P, which very intuitively seen in this figure: apart from the inclination and self-crossings, moving on the surface is exactly like moving on P. For better 3D perception, see the animated version available in `http://www.cs.helsinki.fi/u/totalvit/ksp/coveringspace/`.

.

We will prove that $\gamma \mapsto \gamma(|\gamma|)$ is a covering map as a theorem, but to keep the proof simple we first formulate a lemma which helps in proving the local uniqueness of the paths with moved endpoints, encapsulating the most tedious part of the proof.

Lemma 3.12. *There exists $\epsilon > 0$ that depends only on the global domain P and point s such that for any point $x \in P$ and distinct locally shortest paths $\alpha, \beta \in \mathcal{L}_{sx}$ it holds that $d(\alpha, \beta) \geq \epsilon$.*

Proof. We will prove that the following choice of ϵ works:

$$\epsilon = \frac{1}{3} \min_{\substack{a,b,c \in V \cup \{s\} \\ a \neq b \neq c \neq a}} D_{ab}(c),$$

where $D_{ab}(c)$ denotes the distance from c to line ab. By the nondegeneracy assumptions of Global definition 2.2 and the finiteness of $V \cup \{s\}$, we know that $\epsilon > 0$. This definition also ensures that the distance between any two points in $V \cup \{s\}$ is at least 3ϵ.

We will prove a generalization of the claim, where we only require that paths α and β are in $\mathcal{L}_{px}$ for some $p \in V \cup \{s\}$. Let us write $\alpha = [u_1, \ldots, u_m]$ and $\beta = [v_1, \ldots, v_n]$ such that m and n are as small as possible. Thus $u_1 = v_1 = p \in V \cup \{s\}$, $u_m = v_n = x$ and by the minimality of m and n, it holds that $u_2, \ldots, u_{m-1}, v_2, \ldots, v_{n-1} \in V$. Furthermore, the lists do not contain repeated points or three consecutive collinear points. Without loss of generality we may assume that $m \leq n$. Let us prove the claim by induction on m, by proving the contrapositive form of the claim: if $d(\alpha, \beta) < \epsilon$, then $\alpha = \beta$.

Base case $m = 1$. Now $\alpha = [p]$. If $\beta \neq [p]$, then β visits some point in $(V \cup \{s\}) \setminus \{p\}$, which would give the contradiction $d(\alpha, \beta) \geq 3\epsilon > \epsilon$. Thus $\alpha = \beta$.

Base case $m = 2$. Now $\alpha = [p, x]$. If $n = 2$, β is also a segment, and because $\beta \in \mathcal{L}_{px}$, $\beta = [p, x] = \alpha$. Thus we may assume that $n \geq 3$. Now by the alternative definition of d given in Lemma 3.2, there exists an increasing bijection $f : [0, |\beta|] \to [0, |\alpha|]$ such that for all $t \in [0, |\beta|]$, $||\alpha(f(t)) - \beta(t)|| < \epsilon$. Because $\alpha = [p, x]$, this yields that $D_{px}(\beta(t))$, that is, the distance from $\beta(t)$ to line px, is smaller than ϵ.

Let us rotate and translate our coordinate system to a more convenient one by defining the $\langle w, z \rangle$ coordinate system as shown in Figure 3.7 such that the coordinates for p and x are $\langle 0, 0 \rangle$ and $\langle ||x - p||, 0 \rangle$, respectively. We use w and z as the coordinate functions: $w(\langle a, b \rangle) = a$ and $z(\langle a, b \rangle) = b$.

Choose parameters $0 = t_1 < t_2 < \ldots < t_n = |\beta|$ such that $\beta(t_k) = v_k$ for all $k \in \{1, 2, \ldots, n\}$. Now by the choice of f we know that $||\alpha(f(t_k)) - v_k|| < \epsilon$, and therefore also $|w(\alpha(f(t_k))) - w(v_k)| < \epsilon$. Because $\alpha = [p, x]$, we can rewrite this as

$$|f(t_k) - w(v_k)| < \epsilon.$$

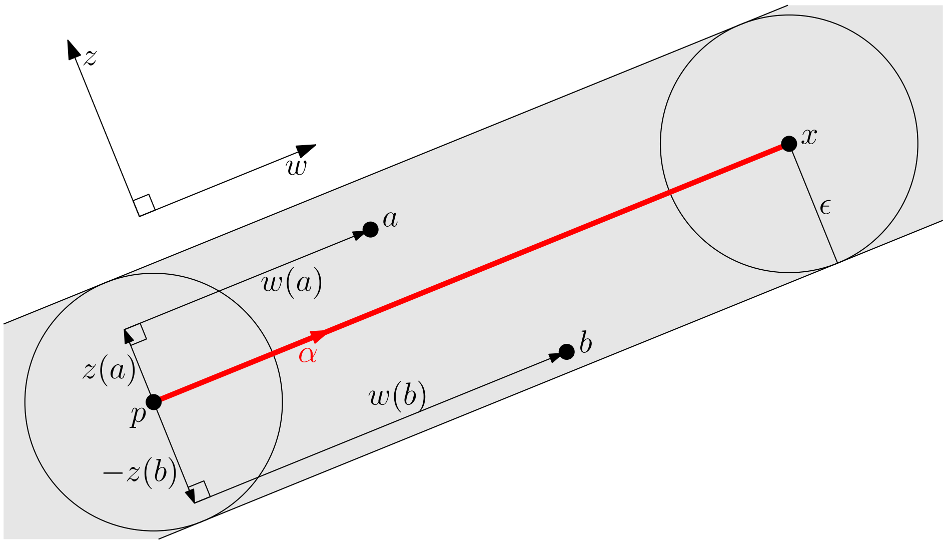


Figure 3.7: To study the base case $m = 2$ of the proof of Lemma 3.12 where $\alpha = [p, x]$, we use a translated and rotated coordinate system $\langle w, z\rangle$ such that p is in origin $\langle 0, 0\rangle$, and x is on the positive w-axis at point $\langle ||x-p||, 0\rangle$. In the figure we illustrate the meaning of coordinates $\langle w(a), z(a)\rangle$ and $\langle w(b), z(b)\rangle$ for example points a and b. Because we know that every point of β has distance less than ϵ to the line px, we know that they lie in the shaded region, that is, the region of points x with coordinate $z(x)$ in range $(-\epsilon, \epsilon)$.

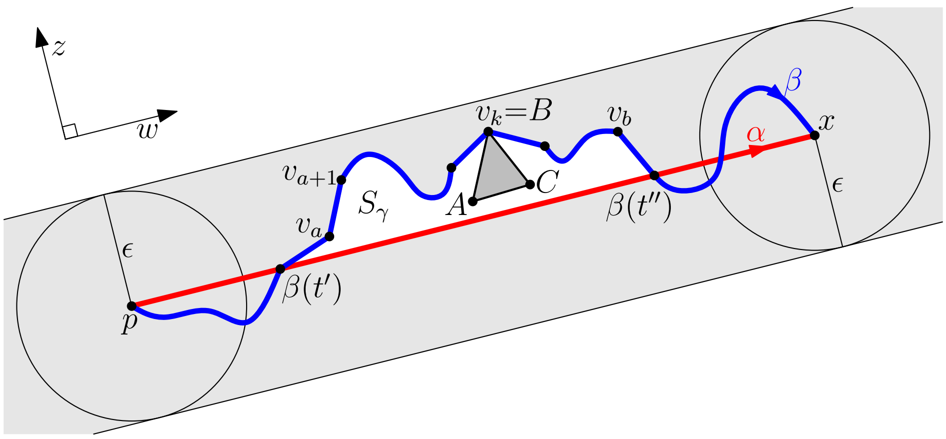


Figure 3.8: In the proof of the base case $m = 2$ in Lemma 3.12, we find a simple polygon S_γ (in white) bounded by a segment $[\beta(t'), \beta(t'')]$ and a subpath $\beta_{[t',t'']} = [\beta(t'), v_a, v_{a+1}, \ldots, v_b, \beta(t'')]$ of β. There exists an internal vertex of the chain v_a such that the internal angle is less than 180°, and because β is a locally shortest path, this means that there exists an obstacle inside S_γ. This leads to a contradiction with the choice of ϵ.

First we prove that the path β is not self-intersecting. We prove this by proving that β is increasing in coordinate w apart from the last vertex, meaning that $w(v_1) < w(v_2) < \ldots < w(v_{n-1})$, and that this actually holds with a margin of 2ϵ: for all $k \in \{1, 2, \ldots, k-2\}$, $w(v_{k+1}) - w(v_k) \geq 2\epsilon$. Using the inequality $|f(t_k) - w(v_k)| < \epsilon$ we can bound

$$\begin{aligned} w(v_{k+1}) - w(v_k) &\geq w(v_{k+1}) - f(t_{k+1}) - w(v_k) + f(t_k) + f(t_{k+1}) - f(t_k) \\ &\geq -|w(v_{k+1}) - f(t_{k+1})| - |w(v_k) - f(t_k)| + f(t_{k+1}) - f(t_k) \\ &> f(t_{k+1}) - f(t_k) - 2\epsilon. \end{aligned}$$

Because f is increasing, $f(t_{k+1}) > f(t_k)$, which now yields the inequality $w(v_{k+1}) - w(v_k) > -2\epsilon$. Now assume that our claim does not hold, i.e. $w(v_{k+1}) - w(v_k) < 2\epsilon$. Combining the two inequalities, we get that $|w(v_{k+1}) - w(v_k)| < 2\epsilon$. Because $z(v_k), z(v_{k+1}) \in (-\epsilon, \epsilon)$, we can now use the Pythagorean theorem to estimate

$$\begin{aligned} ||v_{k+1} - v_k|| &= \sqrt{(w(v_{k+1}) - w(v_k))^2 + (z(v_{k+1}) - z(v_k))^2} \\ &< \sqrt{4\epsilon^2 + 4\epsilon^2} = 2\sqrt{2}\epsilon < 3\epsilon, \end{aligned}$$

which is a contradiction with the choice of ϵ, because $v_k, v_{k+1} \in V$. Therefore the claim $w(v_{k+1}) - w(v_k) \geq 2\epsilon$ holds for all $k \in \{1, 2, \ldots, k-2\}$. Now as $[v_1, v_2, \ldots, v_{n-1}]$ is w-monotone and therefore not self-intersecting, β can self-intersect only if $[v_{n-2}, v_{n-1}]$ intersects with $[v_{n-1}, v_n]$ outside the point v_{n-1} (it cannot intersect with the other part $[v_1, \ldots, v_{n-2}]$ because $w(v_k) \leq w(v_{n-2}) \leq w(v_{n-1}) - 2\epsilon < |\alpha| - \epsilon$ for all $k \in 1, \ldots, n-2$). As the segments share an endpoint, this would mean $\angle(\beta, v_{n-1}) = 0°$ which is a contradiction with the assumption that $\beta \in \mathcal{L}_{px}$. Thus we finally get that β is not self-intersecting.

Now if $\beta \neq [p, x]$, then there exists $s \in [0, |\beta|]$ such that $z(\beta(s)) \neq 0$. Let $t' \in [0, s)$ and $t'' \in (s, |\beta|]$ be the largest and smallest t such that $z(\beta(t)) = 0$. These exists because $z(\beta(0)) = z(p) = 0$, $z(\beta(|\beta|)) = z(x) = 0$ and the preimage of the closed set $\{0\}$ in the continuous function z is compact. By this definition, for all $t' < t < t''$ it holds that $z(\beta(t)) \neq 0$. Therefore the closed path $\gamma = \beta_{[t',t'']}[\beta(t''), \beta(t')]$ bounds a simple polygon S_γ with at least 3 vertices (see Figure 3.8). Choose a, b such that

$$\{t_a, t_{a+1}, \ldots, t_b\} = \{k \in \{1, 2, \ldots, n\} \mid t' < t_k < t''\}.$$

Now $\gamma = [\beta(t'), \beta(t_a), \beta(t_{a+1}), \ldots, \beta(t_b), \beta(t''), \beta(t')]$. Denote the corresponding internal angles at the vertices $\beta(t'), v_a, \ldots, v_b, \beta(t'')$ of the simple polygon S_γ by $\angle', \angle_a, \angle_{a+1}, \ldots, \angle_b, \angle''$. We know that for some $a \leq k \leq b$ it holds that $\angle_k < 180°$, because if angles $\angle_a, \angle_{a+1}, \ldots, \angle_b$ are all at least $180°$, we find a contradiction by computing the total sum of angles T and using the formula for T:

$$T := \angle' + \angle_a + \angle_{a+1} + \ldots + \angle_b + \angle'' > 0° + (b - a - 1)180° + 0° = T.$$

Because β makes a reflex turn at v_k (that is, at t_k), γ also makes a reflex turn at v_k, but because the internal angle of the polygon γ at v_k is less than $180°$, the obstacle connected to v_k must lie inside the simple polygon S_γ (see Figure 3.8). This means that there exists at least three vertices $A, B, C \in V$ inside the simple polygon S_γ. Therefore $z(A), z(B), z(C) \in (-\epsilon, \epsilon)$. Now we will prove that the existence of these three vertices is a contradiction with the choice of ϵ. Without loss of generality, we may assume the ordering $w(A) \leq w(B) \leq w(C)$. Now as the triangle ABC is completely contained in the rectangle $\{\langle w, z\rangle \mid w \in [w(A), w(C)], z \in (-\epsilon, \epsilon)\}$, which means that its surface area at most half of the area of the rectangle. Using the standard formula for triangle surface area, we can express this inequality as

$$\frac{1}{2}||C - A||D_{AC}(B) \leq \epsilon(w(C) - w(A)).$$

Because $w(C) - w(A) \leq ||C - A||$, this yields that $D_{AC}(B) \leq 2\epsilon$, which is a contradiction with the assumption that $D_{AC}(B) \geq 3\epsilon$. Therefore the contrary holds: $\beta = [p, x] = \alpha$.

Induction step for $m \geq 3$. Now we know that also $n \geq 3$. Because $d(\alpha, \beta) < \epsilon$, there exists nondecreasing surjections $f : [0, 1] \to [0, |\alpha|]$ and $g : [0, 1] \to [0, |\beta|]$ such that for all $t \in [0, 1]$,

$$||\beta(g(t)) - \alpha(f(t))|| < \epsilon.$$

Define A as the smallest $t \in [0, 1]$ such that $f(t) = ||u_2 - p||$ or $g(t) = ||v_2 - p||$. The minimum exists because f and g are continuous, and thus the union of their preimages is compact. If $f(A) = ||u_2 - p||$, then $g(A) \leq ||v_2 - p||$, and we get that

$$||\beta(g(A)) - \alpha(f(A))|| = ||[p, v_2](g(A)) - u_2|| < \epsilon,$$

which implies that point u_2 has distance to line pv_2 less than ϵ. As we chose ϵ such that no vertex has distance less than 3ϵ from a line between two other vertices, this yields that $u_2 = v_2$. With a symmetric computation, we get the same result in the other case when $f(A) \leq ||u_2 - p||$ and $g(A) = ||v_2 - p||$. Now α and β have a common prefix $[p, q]$, where q denotes the vertex $u_2 = v_2$. If we denote $L = ||q - p||$, we can write $\alpha = [p, q]\alpha_{[L,|\alpha|]}$ and $\beta = [p, q]\beta_{[L,|\beta|]}$.

By the definition of A we know that $f(A) \leq L$ and $g(A) \leq L$ and that equality holds at least in one of the inequalities. Now as

$$||\beta(g(A)) - \alpha(f(A))|| = ||[p, q](g(A)) - [p, q](f(A))|| < \epsilon,$$

we know that $|f(A) - g(A)| < \epsilon$, so both $f(A)$ and $g(A)$ lie on interval $(L - \epsilon, L]$.

Now we divide the proof into cases, and in each of them we get the result $\alpha = \beta$ either directly or by proving that $d(\alpha_{[L,|\alpha|]}, \beta_{[L,|\beta|]}) < \epsilon$, which also yields the result $\alpha = \beta$ because we can apply the claim for $\mathcal{L}_{qx}$-paths $\alpha_{[L,|\alpha|]}$ and $\beta_{[L,|\beta|]}$ (decreasing m by one) and obtain that they are equal.

- If $||v_3 - q|| < 2\epsilon$: Because any two distinct vertices of $V \cup \{s\}$ have distance at least 3ϵ and $q \in V$, we get that $v_3 \notin V \cup \{s\}$. Therefore $v_3 = x$, and $n = 3$. Since $m \leq n$, $m = 3$, and therefore $\alpha = \beta = [p, q, x]$.

- If $||u_3 - q|| < 2\epsilon$ and $||v_3 - q|| \geq 2\epsilon$: Now we know that $u_3 = x$ and therefore $\alpha = [p, q, x]$. Furthermore, because $v_3 \neq x$, it must hold that $v_3 \in V$, which by our choice of ϵ means that actually $||v_3 - q|| \geq 3\epsilon$. Let $t \in [A, 1]$ such that $g(t) = |[p, q, v_3]|$. Now because

$$L - \epsilon < f(A) \leq f(t) \leq |\alpha| < L + 2\epsilon,$$

we know that $|f(t) - L| < 2\epsilon$. Since α is 1-Lipschitz, by the triangle inequality we get the contradiction

$$\begin{aligned} ||\beta(g(t)) - \alpha(f(t))|| &= ||v_3 - \alpha(f(t))|| \\ &\geq ||v_3 - \alpha(L)|| - ||\alpha(L) - \alpha(f(T))|| \\ &\geq ||v_3 - q|| - |f(T) - L| \\ &> 3\epsilon - 2\epsilon = \epsilon. \end{aligned}$$

- If $||u_3 - q|| \geq 2\epsilon$ and $||v_3 - q|| \geq 2\epsilon$: As we did with A, define B as the smallest $t \in [0, 1]$ such that $f(t) = |[p, q, u_3]|$ or $g(t) = |[p, q, v_3]|$. Now $B > A$ because f and g are non-decreasing.

If $f(B) \leq L$, then $f(B)$ lies on interval $(L - \epsilon, L]$, and therefore $||\alpha(f(B)) - q|| < \epsilon$. Now as $f(B) < |[p, q, u_3]|$, $\beta(g(B)) = v_3$. By combining these results with the assumption $||v_3 - q|| \geq 2\epsilon$ using the triangle inequality, we get that

$$\begin{aligned} ||\beta(g(B)) - \alpha(f(B))|| &= ||v_3 - \alpha(f(B))|| \\ &\geq ||v_3 - q|| - ||\alpha(f(B)) - q|| \\ &> 2\epsilon - \epsilon = \epsilon, \end{aligned}$$

which is a contradiction. Similarly we get a contradiction if $g(B) \leq L$. Thus $f(B) \in (L, L + ||u_3 - q||]$ and $g(B) \in (L, L + ||v_3 - q||]$.

Now we are ready to prove that $d(\alpha_{[L,|\alpha|]}, \beta_{[L,|\beta|]}) < \epsilon$. We do this by defining nondecreasing surjections $u : [0, 1] \to [0, |\alpha| - L]$ and $v : [0, 1] \to [0, |\beta| - L]$ by

$$u(t) = \begin{cases} (f(B) - L) \cdot 2t, & \text{if } t \leq 1/2 \\ f(B + (2t - 1)(1 - B)) - L, & \text{if } t > 1/2 \end{cases}$$

$$v(t) = \begin{cases} (g(B) - L) \cdot 2t, & \text{if } t \leq 1/2 \\ g(B + (2t - 1)(1 - B)) - L, & \text{if } t > 1/2. \end{cases}$$

The idea is that if t is in interval $[1/2, 1]$, we get the claim

$$||\beta_{[L,|\beta|]}(v(t)) - \alpha_{[L,|\alpha|]}(u(t))|| < \epsilon$$

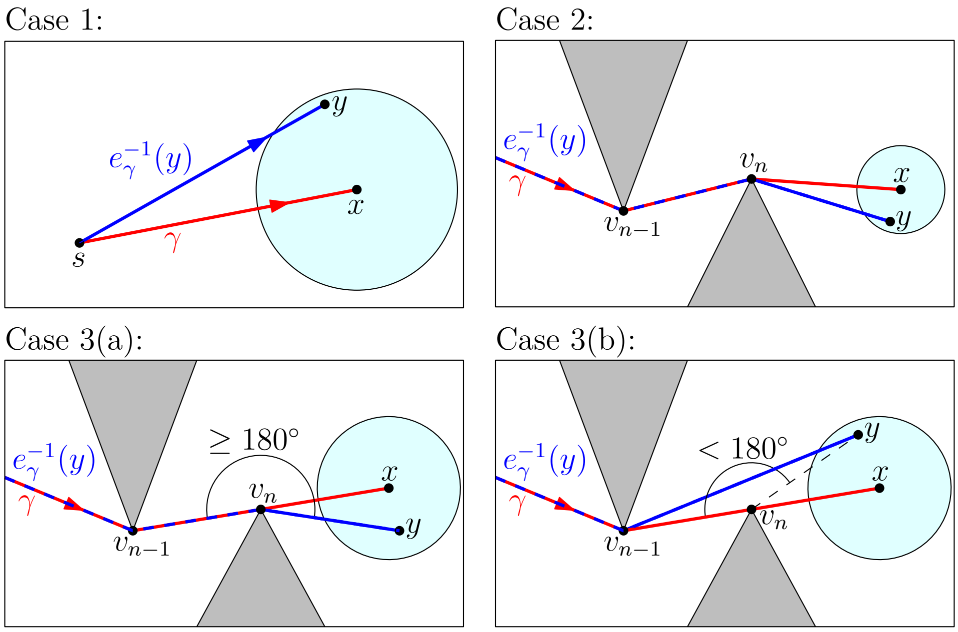


Figure 3.9: Examples of the values of the local inverse function $e_\gamma^{-1}(y)$ of the endpoint function e at point $y \in B(x, r)$ for different γ and y.

---

directly since u and v are just $f - L$ and $g - L$ reparameterized, and as t moves through interval $[0, 1/2]$ we affinely move u and v from q to $\alpha(f(B))$ and $\beta(f(B))$, which keeps the distance small because the endpoints have distance less than ϵ. More exactly, if $t \in [0, \frac{1}{2}]$,

$$\begin{aligned}
&||\beta_{[L,|\beta|]}(v(t)) - \alpha_{[L,|\alpha|]}(u(t))|| \\
&= ||\beta((g(B) - L) \cdot 2t + L) - \alpha((f(B) - L + L) \cdot 2t)|| \\
&= ||[q, v_3]((g(B) - L) \cdot 2t) - [q, u_3]((f(B) - L) \cdot 2t)|| \\
&= ||q + 2t(\beta(g(B)) - q) - q - 2t(\alpha(f(B)) - q)|| \\
&= 2t||\beta(g(B)) - \alpha(f(B))|| < \epsilon.
\end{aligned}$$

Now because $||\beta_{[L,|\beta|]}(v(t)) - \alpha_{[L,|\alpha|]}(u(t))|| < \epsilon$ for all $t \in [0, 1]$, we get that $d(\alpha_{[L,|\alpha|]}, \beta_{[L,|\beta|]}) < \epsilon$.

□

Now we are ready to prove that $\mathcal{L}_s$ is a covering space of P.

Theorem 3.13. *Define the endpoint mapping $e : \mathcal{L}_s \to P$ by*

$$e(\gamma) = \gamma(|\gamma|).$$

This mapping is a covering map. If we use the terminology of Definition 3.11, then we can choose $r_x > 0$ and $\mathcal{U}_x$ for all $x \in P$ such that $\mathcal{U}_x = \{U_\gamma \mid \gamma \in \mathcal{L}_{sx}\}$, where

$$U_\gamma = B_{\mathcal{L}_s}(\gamma, r_x) = \{\alpha \in \mathcal{L}_s \mid d(\gamma, \alpha) < r_x\}.$$

In addition, as for the exact representation of $e|_{U_\gamma}^{-1}$, the following holds:

Let $x \in P$ and $\gamma \in \mathcal{L}_{sx}$. Let $t_1 = 0$ and $0 < t_2 < t_3 < \ldots < t_n \leq |\gamma|$ be the parameters t such that $\gamma(t) \in V$. Denote $v_k = \gamma(t_k)$ for all $k \in \{1, 2, \ldots, n\}$, so now $\gamma = [v_1, v_2, \ldots, v_n, x]$. Denote $e|U_\gamma$ by a shorthand notation e_γ. Now e_γ is an isometry (and therefore also a homeomorphism) $U_\gamma \to B_P(x, r_x)$, $e_\gamma^{-1}(x) = \gamma$, and for all $y \in B_P(x, r_x) \setminus \{x\}$,

1. *If $n = 1$ (implying $\gamma = [s, x]$), $e_\gamma^{-1}(y) = [s, y]$.*
2. *If $n \geq 2$ and points v_{n-1}, v_n and x are not collinear,*
$$e_\gamma^{-1}(y) = \gamma_{[0,t_n]}[v_n, y].$$
3. *If $n \geq 2$ and points v_{n-1}, v_n and x are collinear (including the case $v_n = x$),*
   (a) If $[v_{n-1}, v_n, y]$ makes a reflex turn at v_n,
   $$e_\gamma^{-1}(y) = \gamma_{[0,t_n]}[v_n, y].$$
   (b) Otherwise,
   $$e_\gamma^{-1}(y) = \gamma_{[0,t_{n-1}]}[v_{n-1}, y].$$

See Figure 3.9 for examples of these cases.

Proof. We choose $r_x > 0$ such that for all pairs of points $a, b \in V \cup \{s\}$, if x does not lie on the line ab, the distance from x to line ab is at least r_x. In addition, we require that $r_x < \epsilon/2$, where ϵ is the positive number from Lemma 3.12 such that all distinct paths of $\mathcal{L}_{sx}$ have distance at least ϵ. Such number exists, because $V \cup \{s\}$ is finite.

It is easy to check that for all $\gamma \in \mathcal{L}_{sx}$ and $y \in B_P(x, r_x)$, the explicit paths $e_\gamma^{-1}(y)$ given in the cases of the claim are in U_γ. U_γ does not contain any other paths apart from the ones given explicitly in the cases of the claim, because if U_γ contained two distinct paths $\alpha, \beta \in \mathcal{L}_{sy}$ for some $y \in B_P(x, r_x)$, then by the triangle inequality $d(\alpha, \beta) < 2r_x < \epsilon$, which by Lemma 3.12 cannot happen. Thus e_γ is a bijection $U_\gamma \to B_P(x, r_x)$.

Now it suffices to prove that e_γ is an isometry: If $\alpha, \beta \in U_\gamma$, then $d(\alpha, \beta) = ||e(\alpha) - e(\beta)||$. Clearly by definition $d(\alpha, \beta) \geq ||e(\alpha) - e(\beta)||$. Thus it suffices to prove that for all $p, q \in B_P(x, r_x)$,

$$d(e_\gamma^{-1}(p), e_\gamma^{-1}(q)) \leq ||p - q||.$$

Let us prove this in all cases of the claim separately.

1. Now $e_\gamma^{-1}(p) = [s,p]$ and $e_\gamma^{-1}(q) = [s,q]$, and

$$\begin{aligned} d([s,p],[s,q]) &\leq \max_{t\in[0,1]} ||[s,p]'(t) - [s,q]'(t)|| \\ &= \max_{t\in[0,1]} ||(1-t)s + tp - (1-t)s - tq|| \\ &= \max_{t\in[0,1]} t||p-q|| = ||p-q||. \end{aligned}$$

2. Now there exists a prefix $\alpha \in \mathcal{L}_{sv}$ of γ, where $v \in V$, such that $e_\gamma^{-1}(p) = \alpha[v,p]$ and $e_\gamma^{-1}(q) = \alpha[v,q]$. By Theorem 3.4, we know that

$$\begin{aligned} d(e_\gamma^{-1}(p), e_\gamma^{-1}(q)) &\leq d(\alpha,\alpha) + d([v,p],[v,q]) \\ &= 0 + d([v,p],[v,q]) \leq ||p-q||, \end{aligned}$$

where the last inequality is obtained similarly to case 1.

3. Write $\gamma = [v_1, v_2, \ldots, v_n, x]$ like in the claim of the theorem. If $p = v_n$, $q = v_n$, or both turn angles $\angle([v_{n-1}, v_n, p])$ and $\angle([v_{n-1}, v_n, q])$ are either at most 180° or at least 180°, then the result is obtained similarly to case 2. Otherwise, there exists a point z in the segment pq such that the points v_{n-1}, v_n and z are collinear. Thus we can apply the result of the previous case to get that

$$\begin{aligned} d(e_\gamma^{-1}(p), e_\gamma^{-1}(q)) &\leq d(e_\gamma^{-1}(p), e_\gamma^{-1}(z)) + d(e_\gamma^{-1}(z), e_\gamma^{-1}(q)) \\ &\leq ||p-z|| + ||z-q|| = ||p-q||. \end{aligned}$$

□

3.4 Relation to path homotopy

In this subsection, we will prove that for all $x \in P$, the set of locally shortest paths $\mathcal{L}_{sx}$ is the set of unique shortest paths of each homotopy class of curves from s to x in P, showing that our definition of locally shortest paths is equivalent to that of [4]. We will first briefly introduce the notion of path homotopy and the lift theorem, which allows us to use the covering map properties proven in Theorem 3.13 to reason about homotopy.

Definition 3.14. Let $\alpha, \beta \in \mathcal{C}_{ab}$ for some $a, b \in P$. A continuous function $H : [0,1] \times [0,1] \to P$ is a *homotopy* between the curves α and β if

- $H(s,0) = \alpha'(s)$ for all $s \in [0,1]$,
- $H(s,1) = \beta'(s)$ for all $s \in [0,1]$,
- $H(0,t) = a$ for all $t \in [0,1]$,
- $H(1,t) = b$ for all $t \in [0,1]$.

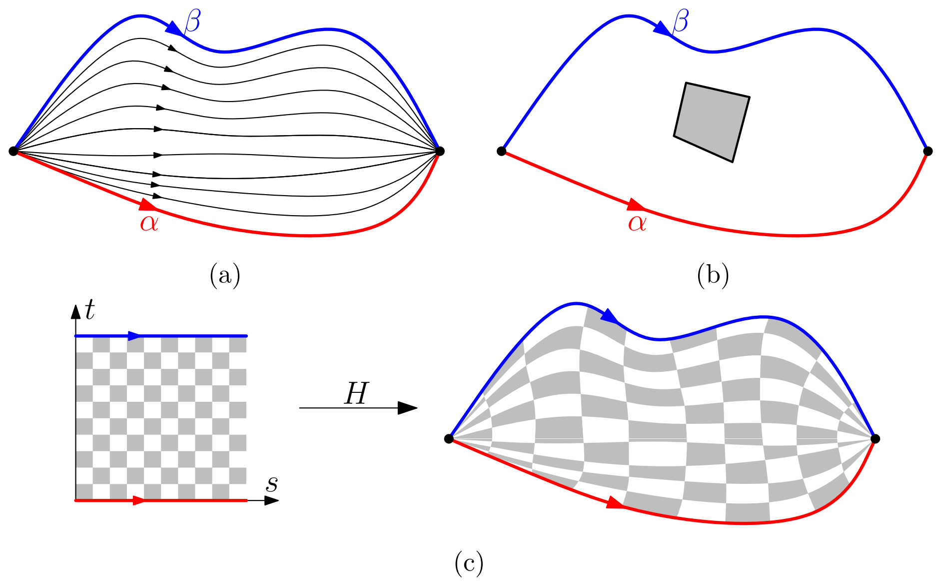


Figure 3.10: α is homotopic to β, if α can be deformed continuously in the domain P to β without moving the endpoints, as illustrated in subfigure (a). More exactly, this deformation is expressed as a continuous function $H : [0,1] \times [0,1] \to P$ like the one in subfigure (c), where the cross section $s \to H(s,t)$ gives the curve α' if $t = 0$ and β' if $t = 1$, and the deformation between the paths when t is between 0 and 1. We also require that $t \to H(0,t)$ and $t \to H(1,t)$ are constant functions to ensure that the endpoints are not moved. Sometimes such a deformation does not exist, as for example in the case of subfigure (b) in which there is an obstacle between the paths.

If there exists a homotopy between α and β, then we say that α and β are *homotopic* curves, and we denote that $\alpha \sim \beta$.

Intuitively, this definition of α and β being homotopic means that we can continuously deform α to β while keeping the endpoints intact: as we increase the time parameter t from 0 to 1, the curve $s \mapsto H(s,t)$ changes continuously from α' to β'. See Figure 3.10 for an illustration of the concept. The following two theorems are basic results of homotopy theory and are presented without proofs. For a more in-depth introduction, see chapter 1 of [7].

Theorem 3.15. *The homotopy relation $\sim$ is an equivalence relation in $\mathcal{C}$ with the following additional properties:*

- *If $\alpha_1, \beta_1 \in \mathcal{C}_{ab}$ and $\alpha_2, \beta_2 \in \mathcal{C}_{bc}$ for some $a, b, c \in P$ such that $\alpha_1 \sim \beta_1$ and $\alpha_2 \sim \beta_2$, then $\alpha_1\alpha_2 \sim \beta_1\beta_2$.*
- *Homotopy is independent of parameterization: if $\gamma : [0, L] \to P$ is*

a curve and $f : [0, x] \to [0, L]$ is a nondecreasing surjection, then $\gamma \circ f \sim \gamma$.

Theorem 3.16. *Let X, Y and Z be metric spaces and $p : X \to Y$ be a covering map. For any continuous mapping $f : Z \to Y$ we say that a continuous mapping $g : Z \to X$ is a* lift *of f if $p \circ g = f$, meaning that the following diagram commutes:*

$$\begin{tikzcd} & X \arrow[d, "p"] \\ Z \arrow[ur, "g"] \arrow[r, "f"'] & Y \end{tikzcd}$$

Now if $x \in X$, $Z = [0, 1]$ or $Z = [0, 1] \times [0, 1]$ and f is a continuous mapping $Z \to Y$ such that $f(0) = p(x)$, then there exists exactly one lift $\tilde{f}$ of f such that $\tilde{f}(0) = x$.

Next we will prove that using the endpoint covering map $e : \mathcal{L}_s \to P$ we can for any curve $\gamma : [0, 1] \to P$ obtain a locally shortest path homotopic to γ by taking the endpoint $\tilde{\gamma}(1)$ of the e-lift $\tilde{\gamma}$ of γ. See Figure 3.11 for some intuition on what this looks like.

Theorem 3.17. *Let $\gamma : [0, 1] \to P$ be a curve with $\gamma(0) = s$. Denote $x = \gamma(1)$. Then there exists a unique locally shortest path $\alpha \in \mathcal{L}_{sx}$ homotopic to γ. We call this α the* shortcut *of γ, and the shortcut of a general curve $\gamma : [0, L] \to P$ is the same as the shortcut of its reparameterization γ'.*

Proof. Since the endpoint function $e : \mathcal{L}_s \to P$ is a covering map, by Theorem 3.16 there exists a lift $\tilde{\gamma}$ such that $\tilde{\gamma}(0) = [s]$. We will prove that $\alpha = \tilde{\gamma}(1)$ is homotopic to γ, and that it is the only path in $\mathcal{L}_{sx}$ homotopic to γ.
Homotopy. We do this by defining

$$X = \{t \in [0, 1] \mid \gamma \sim \tilde{\gamma}(t)\gamma_{[t,1]}\}.$$

Now clearly $0 \in X$, and to prove the claim, we only need to show that $1 \in X$.

Let $p = \sup X \in [0, 1]$. Denote $x = \gamma(p)$. To prove the claim $1 \in X$, it suffices to prove that for some neighborhood I of x in $[0, 1]$ it holds that $I \subset X$, because then we know that $p \in X$ and $p = 1$ (because otherwise I contains larger elements than p).

Because γ and $\tilde{\gamma}$ are continuous, there exists $\delta > 0$ such that for all $t \in [0, 1] \cap (p - \delta, p + \delta) =: I$ it holds that $||\gamma(t) - \gamma(p)|| < r_x$ and $\tilde{\gamma}(t) \in U_{\tilde{\gamma}(p)}$ where r_x and $U_{\tilde{\gamma}(p)}$ are as defined in Theorem 3.13. Note that now for all $t \in I$,

$$\tilde{\gamma}(t) = e^{-1}_{\tilde{\gamma}(p)}(\gamma(t)),$$

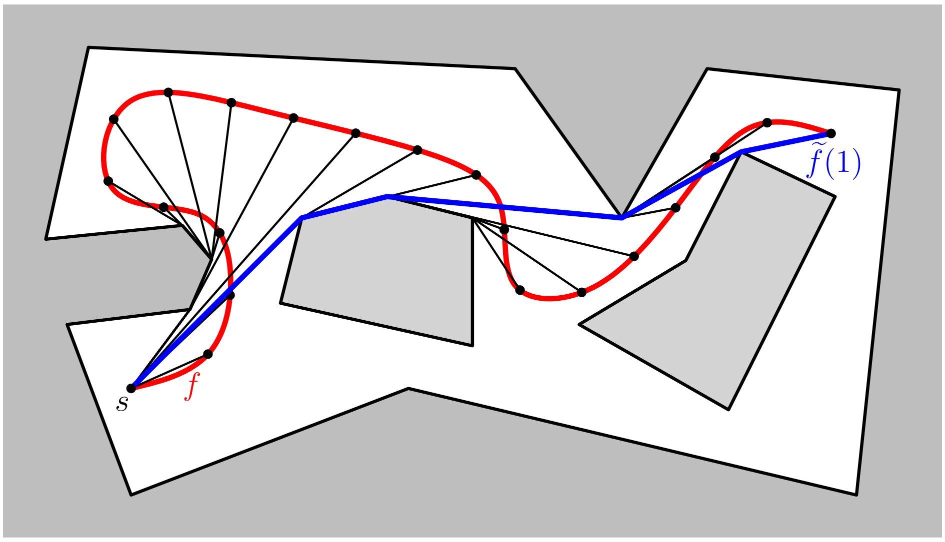


Figure 3.11: By the lift Theorem 3.16 we know that for any curve $\gamma : [0,1] \to P$ with $\gamma(0) = s = e([s])$ there exists a unique lift $\tilde{\gamma}$ of γ such that $\tilde{\gamma}(0) = [s]$, that is, a continuous mapping $[0,1] \to \mathcal{L}_s$ such that $p \circ \tilde{\gamma} = \gamma$. Since $\tilde{\gamma}$ is continuous, as t advances from 0 to 1, the path $\tilde{\gamma}(t)$ changes continuously. Furthermore, because $p \circ \tilde{\gamma}(t) = \gamma(t)$, the endpoint of $\tilde{\gamma}(t)$ follows the curve γ. Intuitively what we do is that we take an elastic string, fix one of its endpoints to s and traverse the path γ with the other endpoint, keeping the string tight at all times. Thus the path $\tilde{\gamma}(1)$ we get in the end follows the same route around the obstacles as the curve γ, and therefore it is intuitive that it is homotopic to γ. We prove this in Theorem 3.17. See the animated version of this figure in in `http://www.cs.helsinki.fi/u/totalvit/ksp/shortcut.gif`.

---

where $e_{\tilde{\gamma}(p)}$ is the local homeomorphism $U_{\tilde{\gamma}(p)} \to B_P(x, r_x)$. By considering all the cases of Theorem 3.13 we see that there exists $v \in V$, a continuous function $g : I \to P$ and a path $\beta \in \mathcal{L}_{sv}$ such that for all $t \in I$,

$$\tilde{\gamma}(t) = \beta[v, g(t), \gamma(t)].$$

Let $a, b \in I$ such that $a < b$. We prove that $\tilde{\gamma}(a)\gamma_{[a,1]} \sim \tilde{\gamma}(b)\gamma_{[b,1]}$ by explicitly constructing the homotopy morphing through paths $\tilde{\gamma}(t)\gamma_{[t,1]}$ as t advances from a to b. More exactly, define $H : [0,1] \times [0,1] \to P$ by defining for all $z \in [0,1]$ and $t \in [a,b]$

$$H\left(z, \frac{t-a}{b-a}\right) = \begin{cases} \beta'(4z) & \text{if } 0 \quad \leq z < 1/4 \\ [v, g(t)]'(4z-1) & \text{if } 1/4 \leq z < 1/2 \\ [g(t), \gamma(t)]'(4z-2) & \text{if } 1/2 \leq z < 3/4 \\ \gamma'_{[t,1]}(4z-3) & \text{if } 3/4 \leq z \leq 1 \end{cases}$$

Now clearly this function is continuous, $H(0,t) = s$ and $H(1,t) = x$ for all $t \in$

$[0,1]$, and as $z \mapsto H(z,0)$ is a reparameterization of $\tilde{\gamma}(a)\gamma_{[a,1]}$ and $z \mapsto H(z,1)$ is a reparameterization of $\tilde{\gamma}(b)\gamma_{[b,1]}$, we have proved that $\gamma(a)\gamma_{[a,1]} \sim \gamma(b)\gamma_{[b,1]}$ for all pairs $a, b \in I$. This means that either $I \subset X$ or $I \cap X = \emptyset$. However, by the definition of supremum, any neighborhood of the supremum of X contains elements of X, and therefore $I \subset X$.

Uniqueness. Let $\beta \in \mathcal{L}_{sx}$ such that $\beta \sim \gamma$. We only need to prove that $\beta = \alpha$. Now there exists a continuous function $H : [0,1] \times [0,1] \to P$ such that $H(0,z) = s$, $H(1,z) = x$, $H(z,0) = \gamma(z)$ and $H(z,1) = \beta'(z)$ for all $z \in [0,1]$. By Theorem 3.16, there exists a unique e-lift $\tilde{H} : [0,1] \times [0,1] \to \mathcal{L}_s$ of H such that $\tilde{H}(0,0) = [s]$.

Define for all $t \in [0,1]$ the functions

- $T_t : [0,1] \to P$ by $T_t(z) = H(z,t)$,
- $\tilde{T}_t : [0,1] \to \mathcal{L}_s$ by $\tilde{T}_t(z) = \tilde{H}(z,t)$,

and for all $z \in [0,1]$ the functions

- $Z_z : [0,1] \to P$ by $Z_z(t) = H(z,t)$,
- $\tilde{Z}_z : [0,1] \to \mathcal{L}_s$ by $\tilde{Z}_z(t) = \tilde{H}(z,t)$.

Now $T_0 = \gamma$, $T_1 = \beta'$, $Z_0 \equiv s$ and $Z_1 \equiv x$. By the uniqueness of lifts shown in Theorem 3.16, we know that $\tilde{T}_0 = \tilde{\gamma}$ and $\tilde{Z}_0 \equiv [s]$. Now that we know that $\tilde{H}(1,0) = \tilde{\gamma}(1) = \alpha$ and $\tilde{H}(0,1) = [s]$, we can use the uniqueness of lifts to get that $\tilde{Z}_1 \equiv \alpha$ and $\tilde{T}_1 = \tilde{\beta}'$, where $\tilde{\beta}'$ is the unique lift of β' such that $\tilde{\beta}'(0) = [s]$.

It also holds that $\tilde{\beta}'(z) = \beta_{[0,z|\beta|]}$ for all $z \in [0,1]$ by the uniqueness of lifts, because $\tilde{\beta}'(0) = \beta_{[0,0]} = [s]$, $\beta_{[0,z|\beta|]} \in \mathcal{L}_s$ and $e(\beta_{[0,z|\beta|]}) = \beta'(z)$ for all $z \in [0,1]$. Now we can compute $\tilde{\beta}'(1)$ in two ways: $\tilde{\beta}'(1) = Z_1(1) = \alpha$ and $\tilde{\beta}'(1) = \beta_{[0,1|\beta|]} = \beta$. Thus $\alpha = \beta$, and we have proved the claim. □

Now we have proved that every homotopy type contains exactly one locally shortest path between given points, that is, the shortcut of all the paths in the homotopy type. There is still a small detail we have not proved: that if we shortcut a path, the shortcut is shorter than the original. This will prove that the locally shortest paths are the shortest paths of their own homotopy types.

Theorem 3.18. *Let $x \in P$ and $\gamma \in \mathcal{P}_{sx}$. Let α be the shortcut of γ. Then $|\alpha| \leq |\gamma|$,*

Proof. Let us reparameterize γ to the unit interval, i.e. $\gamma : [0,1] \to P$ is a curve of $\mathcal{C}_{sx}$.

By the proof of the previous theorem, we know that $\alpha = \tilde{\gamma}(1)$. Define function $f : [0,1] \to \mathbb{R}$ by

$$f(t) = |\tilde{\gamma}(t)\gamma_{[t|\gamma|,|\gamma|]}| = |\tilde{\gamma}(t)| + (1-t)|\gamma|.$$

Clearly $f(0) = |\gamma|$, and the claim is that $f(1) \leq |\gamma|$. Assume the contrary: $f(1) > |\gamma|$. Define

$$p = \min\{t \in [0,1] \mid f(t) = f(1)\}.$$

The minimum exists, because f is continuous. Since $f(0) < f(1)$, it holds that $p > 0$. Denote $x = \gamma(p)$. Let $0 < \epsilon < \min\{p, r_x\}$, where r_x is as defined in Theorem 3.13. Define curves $\beta_1 = \tilde{\gamma}(p-\epsilon)\gamma_{[(p-\epsilon),p]}$ and $|\beta_2| = \tilde{\gamma}(p)$. Now $\beta_1, \beta_2 \in \mathcal{C}_{s\gamma(p)}$, $|\beta_1| = f(p-\epsilon) - (1-p)|\gamma|$ and $|\beta_2| = f(p) - (1-p)|\gamma|$, and since $f(p-\epsilon) < f(p)$, $|\beta_1| < |\beta_2|$. Furthermore, using Theorem 3.4 we may estimate

$$d(\beta_1, \beta_2) \leq d(\tilde{\gamma}(p-\epsilon), \tilde{\gamma}(p)) + d(\gamma_{[(p-\epsilon)|\gamma|, p|\gamma|]}, [\gamma(p)]).$$

Now $\tilde{\gamma}(p-\epsilon) = e^{-1}_{\tilde{\gamma}(p)}(\gamma(p-\epsilon))$ and $\tilde{\gamma}(p) = e^{-1}_{\tilde{\gamma}(p)}(\gamma(p))$, where $e_{\tilde{\gamma}(p)}$ is as defined in Theorem 3.13. Because $e_{\tilde{\gamma}(p)}$ is an isometry and γ is $|\gamma|$-Lipschitz,

$$d(\tilde{\gamma}(p-\epsilon), \tilde{\gamma}(p)) \leq ||\gamma(p) - \gamma(p-\epsilon)|| \leq |\gamma|\epsilon.$$

The same bound holds for the second term also, because

$$d(\gamma_{[(p-\epsilon)|\gamma|, p|\gamma|]}, [\gamma(p)]) \leq \max_{t \in [p-\epsilon, p]} ||\gamma(t) - \gamma(p)|| \leq \epsilon.$$

Therefore $d(\beta_1, \beta_2) \leq 2\epsilon$. Summing up, for every sufficiently small $\epsilon > 0$ we found a path β_1 that is shorter than $\beta_2 = \tilde{\gamma}(p)$ and with distance at most 2ϵ from $\tilde{\gamma}(p)$, which is a contradiction with $\tilde{\gamma}(p)$ being a locally shortest path. $\square$

Shortcuts also give a convenient way to consider local modifications to paths, as we can move an endpoint of a locally shortest path simply by taking the shortcut of the path concatenated with a short segment.

Theorem 3.19. *Define for all $a, b \in P$ such that the segment $[a,b]$ is contained in P a function $M_{ab} : \mathcal{L}_{sa} \to \mathcal{L}_{sb}$ such that $M_{ab}(\gamma)$ is the shortcut of $\gamma[a,b]$. Now M_{ab} is bijective, M_{ba} is the inverse of M_{ab}, and M_{ab} changes the length of a path by at most $||b-a||$, i.e. for all $\gamma \in \mathcal{L}_{sa}$,*

$$\Big| |M_{ab}(\gamma)| - |\gamma| \Big| \leq ||b-a||.$$

Proof. For all $\gamma \in \mathcal{L}_{sa}$, $M_{ab}(\gamma)$ is homotopic to $\gamma[a,b]$, which means that $M_{ba}(M_{ab}(\gamma))$ is homotopic to $\gamma[a,b][b,a]$. Because $\gamma[a,b][b,a]$ is homotopic to γ, and both γ and $M_{ba}(M_{ab}(\gamma))$ are paths of $\mathcal{L}_{sa}$, by the uniqueness result of Theorem 3.17 we know that $\gamma = M_{ba}(M_{ab}(\gamma))$. Thus M_{ba} is the inverse of M_{ab}.

If $\gamma \in \mathcal{L}_{sa}$, then since $M_{ab}(\gamma)$ is the shortcut of $\gamma[a, b]$, by Theorem 3.18 it holds that

$$|M_{ab}(\gamma)| \leq |\gamma[a, b]| = |\gamma| + ||b - a||.$$

Similarly, because γ is the shortcut of $M_{ab}(\gamma)[b, a]$,

$$|\gamma| \leq |M_{ab}(\gamma)[b, a]| = |M_{ab}(\gamma)| + ||b - a||.$$

By combining these inequalities, we get the claim

$$\Big||M_{ab}(\gamma)| - |\gamma|\Big| \leq ||b - a||.$$

□

As we have now formed a one-to-one correspondence between locally shortest paths and homotopy types, we can now find the size of the set $\mathcal{L}_{sx}$ for any $x \in P$ by the size of the homotopy group of the domain.

Theorem 3.20. *If $x \in P$, then*

$$|\mathcal{L}_{sx}| = \begin{cases} 1, & \text{if } P \text{ is simply connected} \\ \infty, & \text{if } P \text{ is not simply connected.} \end{cases}$$

Thus if P contains a hole, $|\mathcal{L}_{sx}| = \infty$.

4 Path sorting

We will now develop a method for finding kth shortest paths for all $k \in \mathbb{Z}_+$ in $\mathcal{L}_{sg}$ by reducing the problem to the problem of listing paths in a weighted graph ordered by length, and using existing graph algorithms to solve it. This is possible because we proved in 3.9 that all locally shortest paths are polygonal chains through polygon vertices, and thus can be expressed as paths in a graph where the set of vertices is $V \cup \{s, g\}$. This algorithm idea was briefly described in [4]. Throughout this section, we will assume that $|\mathcal{L}_{sg}| = \infty$, which by Theorem 3.20 is equivalent to P being not simply connected. The case when P is simply connected simply reduces to finding the shortest path.

4.1 Reduction to a graph problem

Global definition 4.1. The visibility graph of P is the weighted undirected graph $G = (V_G, E_G, w_G)$ such that $V_G = V \cup \{s, g\}$ and $\{a, b\} \subset V_G$ is in the edge set E_G if $a \neq b$ and the segment $[a, b]$ is contained in the domain P, i.e. $\mathrm{Im}[a, b] \subset P$. We define the weight mapping $w_G : E_G \to \mathbb{R}$ for graph G simply as the Euclidean distance between the vertices: $w_G(\{a, b\}) = ||b - a||$. See Figure 1.1 for an example.

Now every path from s to g in the visibility graph G naturally corresponds to a path in $\mathcal{P}_{sg}$. Let $\mathcal{P}'_{sg}$ be the set of paths from s to g in G. We can define this correspondence as the mapping $c : \mathcal{P}'_{sg} \to \mathcal{P}_{sg}$ by

$$c((v_1, v_2, \dots, v_n)) = [v_1, v_2, \dots, v_n].$$

Because of how we weighted the graph G, c preserves path length.

Let us prove that c is injective, i.e. no two paths of $\mathcal{P}_{sg}$ can be written as the c of two distinct paths of $\mathcal{P}'_{sg}$. Assume the contrary: there exists distinct paths $A = (x_1, x_2, \dots, x_m)$ and $B = (y_1, y_2, \dots, y_n)$ such that $A, B \in \mathcal{P}'_{sg}$ but $c(A) = c(B)$. A is not a prefix of B and B is not a prefix of A, because otherwise $|c(A)|$ would not be equal to $|c(B)|$. Thus there exists the smallest $t \in \{2, \dots, \min\{n, m\}\}$ such that $x_t \neq y_t$. Denote $r = x_{t-1} = y_{t-1}$ and $l = ||[x_1, x_2, \dots, x_{t-1}]||$. Without loss of generality we may assume that $||x_t - r|| \geq ||y_t - r||$. Now because $c(A) = c(B)$, $\alpha(l + ||y_t - r||) = y_t$ is on segment $[r, x_t]$, and therefore the three distinct vertices $\{r, x_t, y_t\} \subset V_G$ are collinear, which is a contradiction with the nondegeneracy assumption of the Global definition 2.2.

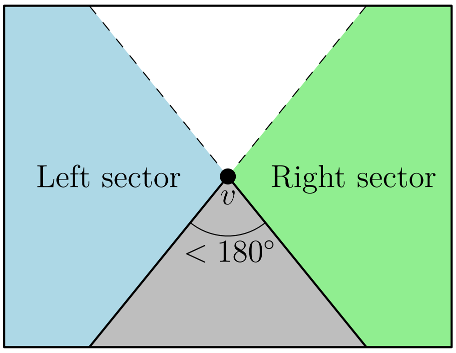
By Theorem 3.9, $\mathcal{L}_{sg}$ is the image $c(\mathcal{L}'_{sg})$, where we define $\mathcal{L}'_{sg}$ as the set of paths $(v_1 = s, v_2, \dots, v_n = g)$ such that $[v_{k-1}, v_k, v_{k+1}]$ makes a reflex turn at v_k for all $k \in \{2, 3, \dots, n-1\}$. This means that $c|_{\mathcal{L}'_{sg}}$ is a bijection $\mathcal{L}'_{sg} \to \mathcal{L}_{sg}$ that preserves path lengths, and the problem of sorting paths in $\mathcal{L}_{sg}$ can be solved by sorting the paths in $\mathcal{L}'_{sg}$. Thus we have now transformed the problem into a problem of finding kth shortest in a constrained set of paths in the visiblity graph.

Computing the visibility graph in polygonal domains is a classical problem in computational geometry. There exist algorithms for computing the graph in $O(|V| \log |V| + |E_G|)$ time [6, 12]. This running time is output sensitive, because it depends on the output size $|E_G|$, which is $O(|V|^2)$, reached for example when all the vertices see each other.

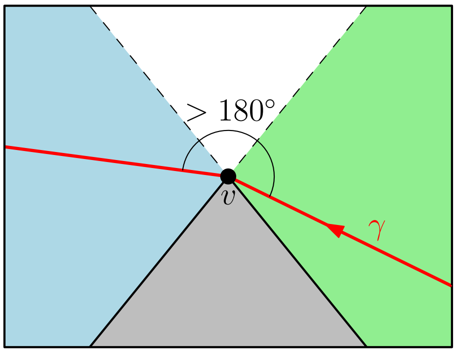
4.2 Graph augmentation

There already exist algorithms for listing paths ordered by the total weight in a weighted directed graph [3]. These algorithms are not immediately applicable to our problem, because our paths have additional constraints: we require that for all three consecutive vertices a, b, c of the path, $[a, b, c]$ makes a reflex turn at b. In this subsection, we reduce the problem further into a problem where we can apply these algorithms.

The restriction means that when the path goes through a vertex, it has to come in through the left sector or the right sector shown in Figure 4.1a, and go out from the opposite sector, turning left if it comes in through the right sector, and vice versa (see Figure 4.1b). So, when traversing paths in the graph, in addition to the information about the node we are currently in, we need to know which direction we are looking at. We will construct a



(a) If the obstacle angle at vertex $v \in V$ is convex, then the left and right sector are the infinite sectors between the extensions of the polygon edges. Otherwise, the sectors are empty.



(b) Locally shortest paths make reflex ($\geq 180°$) turns in all vertices. Therefore they either come in through the right sector and comes out of the left sector, turning left (in the figure), or come in through the left sector and comes out of the right sector, turning right.

Figure 4.1

---

graph where the vertex set is augmented to contain this direction information. More exactly, the vertices will be pairs consisting of the vertex of the original graph G we are currently in, and the direction vector. In order to have a single source and target vertices, we add the s and g to the vertex set.

Our augmented graph is a weighted directed graph $G' = (V_{G'}, E_{G'}, w_{G'})$. We could define the vertex set as

$$\{s, g\} \cup \{(v, d) \mid v \in V_G, d \in \mathbb{R}^2, ||d|| = 1\},$$

that is, vertices are either s, g or some vertex v and the direction being looked at as a unit vector. However, for vertex $v \in V$ we require only directions from/to vertices that see v, so we restrict the set of vertices to a finite subset:

$$\begin{aligned} V_{G'} = \{s, g\} &\cup \{(v, (x - v)/||x - v||) \mid x, v \in V_G, \{x, v\} \in E_G\} \\ &\cup \{(v, (v - x)/||v - x||) \mid x, v \in V_G, \{x, v\} \in E_G\}. \end{aligned}$$

Let us look at what kinds of edges we require. To enable turning, for each $v \in V$ if $(d_1, d_2, \ldots, d_n)$ are the directions $d \in \mathbb{R}^2$ with $||d|| = 1$ such that d points towards the right sector at v ordered from left to right by angle, we add edges

$$E_{\text{right}}(v) = \{((v, d_1), (v, d_2)), ((v, d_2), (v, d_3)), \ldots, ((v, d_{n-1}), (v, d_n))\},$$

We define $E_{\text{left}}(v)$ by swapping words "right" and "left" in the definition.

If we are in a vertex, looking towards another vertex, we should be able to move forward to that vertex, after which we still look towards the same direction. Thus we define the following edges:

$$E_{\text{move}} = \{((a,d),(b,d)) \mid \{a,b\} \in E_G \text{ and } d = (b-a)/||b-a||\}.$$

To allow going to any direction from the source vertex s and coming from any direction to the target vertex g, we add edges to/from all directions from them:

$$E_s = \{(s,(s,d)) \mid (s,d) \in V_{G'}\}$$
$$E_g = \{((g,d),g) \mid (g,d) \in V_{G'}\}.$$

In total, our set of edges will be the disjoint union of these sets:

$$E_{G'} = \bigcup_{v \in V} E_{\text{left}}(v) \cup \bigcup_{v \in V} E_{\text{right}}(v) \cup E_{\text{move}} \cup E_s \cup E_g.$$

See Figure 4.2 for an example of the transform from G to G' in a neighborhood of a vertex. The weight function $w_{G'} : E_{G'} \to \mathbb{R}$ will only charge for moving between vertices, not turning: If $e = ((a,d),(b,d)) \in E_{\text{move}}$,

$$w_{G'}(e) = w_G(\{a,b\}) = ||b-a||,$$

and if $e \in E_{G'} \setminus E_{\text{move}}$, $w_{G'}(e) = 0$.

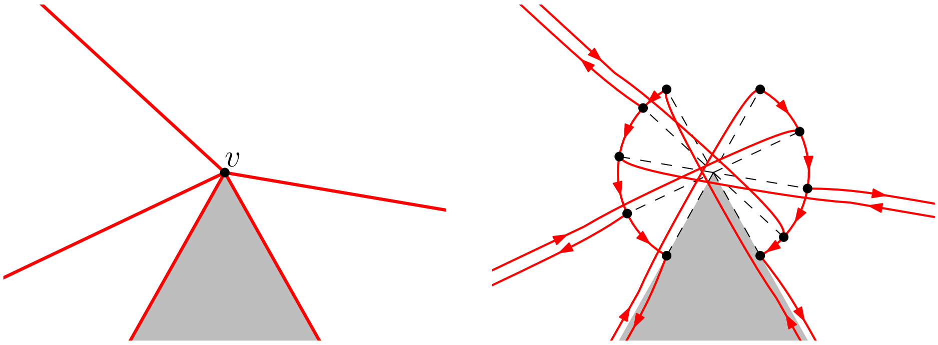
Now by this construction, each directed path from s to g in this graph G' corresponds to a path of $\mathcal{L}'_{sg}$, if we remove the information about directions. Let us define this correspondence $c' : \mathcal{L}''_{sg} \to \mathcal{L}'_{sg}$ more formally. Let $\mathcal{L}''_{sg}$ be the set of directed paths from s to g in G'. Now any path $A \in \mathcal{L}''_{sg}$ can be written as $(s,(v_1,d_1),(v_2,d_2),\ldots,(v_n,d_n),g)$, and necessarily $v_1 = s$ and $d_n = g$. We define $c'(A)$ as the path of corresponding vertices where we remove the duplicates, i.e. $(v_{a_1}, v_{a_2}, \ldots, v_{a_m})$, where $1 = a_1 < a_2 < a_m = n$ and $\{a_2,\ldots,a_m\} = \{k \in \{2,\ldots,n\} \mid v_k \neq v_{k-1}\}$. Directly by the construction, c' defines a bijection $\mathcal{L}''_{sg} \to \mathcal{L}'_{sg}$, and it preserves the length of the path.

Now $c \circ c'$ defines a bijection from the set of directed paths from s to g in G' to $\mathcal{L}_{sg}$ that preserves path lengths, and thus our problem is reduced to listing those directed paths of G'.

It turns out that asymptotically this augmentation does not increase the complexity of the graph: Each edge of E_G adds only at most two vertices to the graph G' (to both of its endpoints), and therefore $|V_{G'}| = O(|E_G|)$. By the construction, each vertex in $V_{G'}$ is adjacent to a constant number of edges in $E_{G'} \setminus E_{\text{move}}$, and there are two edges in E_{move} for each edge in E_G. Thus $|E_{G'}| = O(|E_G|)$.

The graph G' can be constructed easily from the visibility graph G. The only nontrivial operation is the sorting of the vertices (v,d) for all $v \in V$ by the angle of d, which takes $O(|E_{G'}| \log |V|)$ time. Thus the total time complexity of building the augmented graph is

$$O(|V| \log |V| + |E_G| + |E_{G'}| \log |V|) = O(|E_G| \log |V|).$$



(a) (b)

Figure 4.2: Example of the augmentation of the undirected visibility graph G (in (a)) into a directed graph G' (in (b)) in the neighborhood of a vertex $v \in V$. The edges of the graphs are drawn in red. In G', v is replaced with eight vertices in this case, one for each possible direction one can move from v or to v in G. Each edge in G is duplicated into two directed edges between the vertices of appropriate directions. In addition, to allow restricted turning in the vertex, we add the circular edges of weight 0 between the eight vertices shown. It can be seen that any path in G' through the neighborhood of v makes a reflex turn at v.

4.3 Path sorting in the graph

As we have now successfully reduced our problem to sorting directed paths by total weight in a weighted directed graph, the next step is to study the literature on how fast this problem can be solved. Before that, we have to decide exactly how we want to query the list of these paths. Let us define multiple query types, as different applications require different kinds of questions we want to answer.

1. Given $k \in \mathbb{Z}_+$, find k shortest paths of $\mathcal{L}_{sg}$.
2. Given $k \in \mathbb{Z}_+$, find k shortest paths of $\mathcal{L}_{sg}$ ordered by length.
3. Given $k \in \mathbb{Z}_+$, find kth shortest path in the list of paths in $\mathcal{L}_{sg}$ ordered by length.
4. Iterate through all the paths of $\mathcal{L}_{sg}$ ordered by length.

Of course, there might be multiple paths with equal lengths in $\mathcal{L}_{sg}$. We choose to break these ties arbitrarily.

The state-of-the-art algorithm for finding shortest directed paths in weighted graphs between given vertices is given by Eppstein in [3]. For given $k \in \mathbb{Z}_+$, it finds the k shortest directed paths in a directed weighted graph

$X = (V_X, E_X)$ from $a \in V_X$ to $b \in V_X$ in $O(|E_X| + |V_X| \log |V_X| + k)$ time. Therefore, it answers question 1 in time

$$C_1 = O(|E_G| \log |E_G| + k) = O(|E_G| \log |V| + k).$$

However, the algorithm does not find the paths ordered by length. Thus for the second question, we need to sort them in $k \log k$ time, resulting in time complexity

$$C_2 = O(|E_G| \log |V| + k \log k).$$

In the case where we are only interested in the path among the k shortest paths with largest length, we need not sort the whole list of paths, but only iterate through it to find the element with largest length. Thus it takes the same time as question 1.

$$C_3 = C_1 = O(|E_G| \log |V| + k).$$

When iterating the paths ordered by length, we do not know in advance how many shortest paths we want to find, so we cannot apply the algorithm of Eppstein directly. However, we can achieve almost the same time complexity by repeatedly making queries of type 2 with increasing k. We make these queries for $k = 2^0, 2^1, 2^2, \ldots$ as long as required. If in total we expand $k \geq 1$ shortest paths, and $x \geq 0$ is the smallest integer such that $2^x \geq k$, then $2^x = \Theta(k)$, $x = \Theta(\log k)$ and the total time complexity is

$$\begin{aligned} C_4 &= O(x|E_G| \log |V| + \sum_{i=0}^{x} 2^i \log 2^i) \\ &= O(|E_G| \log |V| \log k + \sum_{i=0}^{x} i 2^i) \\ &= O(|E_G| \log |V| \log k + k \log k). \end{aligned}$$

As all of C_1, C_2, C_3 and C_4 are larger than the complexity of building the augmented graph, they give the running time of the whole algorithm including the building time of the visibility graph and the augmented graph. As the worst case bound for $|E_G|$ is $O(|V|^2)$, we get the following running times expressed only using the input size $|V|$ and the parameter k:

$$\begin{aligned} C_1 &= O(|V|^2 \log |V| + k) \\ C_2 &= O(|V|^2 \log |V| + k \log k) \\ C_3 &= O(|V|^2 \log |V| + k) \\ C_4 &= O(|V|^2 \log |V| \log k + k \log k). \end{aligned}$$

Furthermore, the time complexity may depend on the format we want to get the paths in, for example whether we only require the path lengths or the full list of polygon vertices the path goes through. The time complexity $O(|E_X|+|V_X|\log|V_X|+k)$ is achieved in the algorithm of Eppstein when the output for a path is a monoidal product of some properties of the path. More exactly, we define a set S of *properties*, a mapping from edges to properties $f : E_{G'} \to S$ and a binary operator $\otimes : S \times S \to S$ such that $(S, \otimes)$ is a monoid, and we require that both f and $\otimes$ are computable in constant time. The algorithm outputs the monoidal product

$$f(e_1) \otimes f(e_2) \otimes \cdots \otimes f(e_n)$$

for a path consisting of edges $e_1, e_2, \ldots, e_n \in E_{G'}$.

Many properties can be expressed using monoid products of the edge list. Here are some examples of such properties.

- The total length of the path: Define $S = \mathbb{R}$, $f = w_{G'}$ and for all $a, b \in S$, $a \otimes b = a + b$, because then

  $$f(e_1) \otimes f(e_2) \otimes \cdots \otimes f(e_n) = w_{G'}(e_1) + w_{G'}(e_2) + \ldots + w_{G'}(e_n).$$

- The edge list of the path in the augmented graph G' in implicit form: Define S as the set of lists of edges in $E_{G'}$, f as the mapping from an edge to a list of size 1 containing it ($f(e) = (e)$), and $\otimes$ as the concatenation of two lists. Now

  $$f(e_1) \otimes f(e_2) \otimes \cdots \otimes f(e_n) = (e_1) \otimes (e_2) \otimes \cdots \otimes (e_n) = (e_1, e_2, \ldots, e_n).$$

  The naive implementation of $\otimes$ as concatenation of arrays is not constant time, but if we ignore nested lists, we can implement $\otimes$ as $a \otimes b = (a, b)$ in constant time, because it essentially creates a list of size two with pointers to already existing lists. Now for example the implementation may compute

  $$(e_1) \otimes (e_2) \otimes (e_3) \otimes (e_4) = ((e_1), (((e_2), (e_3)), (e_4))),$$

  so the exact result depends on the order in which we compute it. However, if we ignore the nested lists, the result is always equivalent to (e_1, e_2, e_3, e_4). By traversing the tree of the nested list, we can construct the list of edges in time linear in the length of the list.

- The edge list of the path without turning edges, which is the same as the edge list in the corresponding visibility graph G or the list of segments in the corresponding locally shortest path in P: We can modify the previous idea to ignore edges of length zero, i.e. for all $e \in E_{G'}$,

  $$f(e) = \begin{cases} (), & \text{if } w_{G'}(e) = 0 \\ (e), & \text{if } w_{G'}(e) \neq 0. \end{cases}$$

5 kth shortest path map

In this section, our focus shifts to the query version of the problem of finding kth shortest paths from a fixed point s to an arbitrary point x. We do this by constructing the kth shortest path map (k-SPM), a generalization of the shortest path map data structure presented in [10]. The k-SPM is a subdivision of the domain P into cells such that we can read the kth shortest path from s to any given point $x \in P$ by locating the cell of the k-SPM containing the point x. We will prove that this structure is well-behaved enough such that we can preprocess it to support querying kth shortest paths in $O(\log k + \log |V|)$ time using standard point location query algorithms of [2, 11].

Throughout this section, like the previous section, we will only consider the nontrivial case in which P is not simply connected, ensuring that $|\mathcal{L}_{sx}| = \infty$ for all $x \in P$.

5.1 Covering space $\mathcal{L}_s$ in 3D

Before going into the details of defining kth shortest paths, let us try to get some intuitive understanding of the problem. Consider the structure of the set $\mathcal{L}_s$ as a covering space of the domain P as established in Theorem 3.13 with covering map $e(\gamma) = \gamma(|\gamma|)$. This set can be projected as a three-dimensional set using map $D : \mathcal{L}_s \to \mathbb{R}^3$ where path $\gamma \in \mathcal{L}_{sg}$ with endpoint $e(\gamma) = (x, y)$ is projected to point $(x, y, |\gamma|)$. See Figure 3.6 for an example of what the image $D\mathcal{L}_s$ of this projection looks like. Listing the paths of $\mathcal{L}_{sq}$ ordered by length is the same as listing the points in $D\mathcal{L}_s$ with x- and y-coordinates matching q ordered by the z-coordinate. However, D is not necessarily injective, as there might be two distinct paths $\alpha, \beta \in \mathcal{L}_{sq}$ with $|\alpha| = |\beta|$. This can be interpreted as $\operatorname{Im} D$ containing the same point multiple times, more exactly, point $u \in \mathbb{R}^3$ exactly $|D^{-1}\{u\}|$ times. This is finite due to Theorem 3.10.

By Theorem 3.9, we know that all paths $\gamma \in \mathcal{L}_{sq}$ can be written as $[s, q]$ or $\alpha[v, q]$, where $v \in V$, $\alpha \in \mathcal{L}_{sv}$ and γ makes a reflex turn at v, more exactly, $\angle(\gamma, |\gamma| - ||q - v||) \geq 180°$. For $v \in V$ and $\alpha \in \mathcal{L}_{sv}$, let P_α be the set of points $p \in P$ such that $\operatorname{Im}[v, p] \subset P$ and $\gamma[v, p]$ makes a reflex turn at v. Let P_s be set of points $p \in P$ such that $\operatorname{Im}[s, p] \subset P$. It turns out that sets P_α and P_s are always polygons. Now we can rewrite $\mathcal{L}_s$ using these sets

$$\mathcal{L}_s = \{[s, p] \mid p \in P_s\} \cup \bigcup_{v \in V} \bigcup_{\alpha \in \mathcal{L}_{sv}} \{\alpha[v, p] \mid p \in P_\alpha\}.$$

Using this decomposition, we decompose the three-dimensional projection

$D\mathcal{L}_s$ into parts as follows:

$$\begin{aligned} D\mathcal{L}_s &= D\{[s,p] \mid p \in P_s\} \cup \bigcup_{v \in V} \bigcup_{\alpha \in \mathcal{L}_{sv}} D\{\alpha[v,p] \mid p \in P_\alpha\} \\ &= C(s, 0, P_s) \cup \bigcup_{v \in V} \bigcup_{\alpha \in \mathcal{L}_{sv}} C(v, |\alpha|, P_\alpha), \end{aligned}$$

where $C(p, h, \Omega)$ is the upwards cone drawn to point (p_x, p_y, h) and intersected with the set $\Omega \times \mathbb{R}$, i.e.

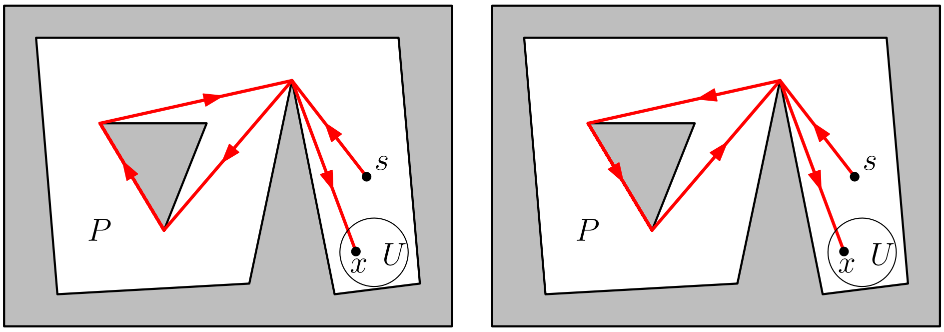
$$C(p, h, \Omega) = \{(x, y, h + ||(x, y) - p||) \mid (x, y) \in \Omega\}.$$

Thus the structure of the set $\mathcal{L}_s$ in 3D is a union of polygonal slices of cones. All these cones are congruent, meaning that they have speed of ascent 1. To get the kth shortest paths from s to all points of P, we need to understand the structure of the kth level of this arrangement of cones.

5.2 kth shortest paths

In section 4, we defined the kth shortest path from s to g simply as the kth element when the paths of $\mathcal{L}_{sg}$ are ordered by length, breaking possible ties in path length arbitrarily. Figure 5.1 shows that it is possible that there exists a neighborhood in the domain such that for all x within the neighborhood, there are two paths in $\mathcal{L}_{sx}$ with equal lengths, and furthermore the configuration is not degenerate, so we cannot plainly disallow such cases. In [4], the problem was averted by stating that it can be worked around by using a perturbation argument. However, we choose to solve the problem more head-on by carefully choosing a tie-breaking scheme that orders the paths consistently also in the neighborhood of x.

To break these kinds of ties, we will refine the ordering of paths by lengths by imposing ordering on some pairs of paths that previously compared equal. To be able to do this, the resulting ordering will be a partial ordering, meaning that some paths will be incomparable. However, it will still compare paths with distinct lengths simply by length. In the case of Figure 5.1, we choose the path that turned less in the last turn before the end of the path to be have smaller length in the refined ordering, so the path of (a) comes before path of (b). Intuitively, this can be visualized by imagining that the boundary is round around the vertex, which means that the path making a larger turn comes out later. In general, we compare the turn angles of the first turn that is different in the paths being compared, counting from the end of the paths (more exactly, the turn at the beginning of the longest common suffix of the paths). This kind of ordering is possible only if the paths have a common part in the end, but it will turn out that these cases are the only ones that require tie-breaking.



(a) (b)

Figure 5.1: The paths of (a) and (b) are both in $\mathcal{L}_{sx}$ for all $x \in U$, and they have equal lengths. This does not change even if we perturb the vertices of the domain, so the configuration is not degenerate.

Definition 5.1. Let $\alpha, \beta \in \mathcal{P}_{ab}$. Let

$$x = \max_{t \in [0,|\alpha|] \cap [0,|\beta|]} \alpha_{t::} = \beta_{t::}.$$

Now we define $\alpha_{x::} = \beta_{x::}$ as the *longest common suffix* of α, and denote it by $\mathrm{LCS}(\alpha, \beta)$.

Well-definedness proof. The maximum exists, because

$$S = \{t \in [0,|\alpha|] \cap [0,|\beta|] \mid \alpha(|\alpha| - t) = \beta(|\beta| - t)\}$$

is a closed set, since it can be written as $f^{-1}\{0\}$, where f is a continuous function $[0,|\alpha|] \cap [0,|\beta|] \to \mathbb{R}$ defined by

$$f(t) = ||\alpha(|\alpha| - t) - \beta(|\beta| - t)\}||.$$

Now $0 \in S$, and x is the largest element of the component of S containing the number 0. □

Definition 5.2. For all $x \in P$, we define a partial order refinement $\preceq_x$ for the ordering of paths in $\mathcal{L}_{sx}$ by the path length such that $\alpha \preceq_x \beta$ if one of the following holds:

- $\alpha = \beta$.
- $|\alpha| < |\beta|$.
- $|\alpha| = |\beta|$, $\alpha \neq \beta$, and paths α and β have nontrivial common suffix, and counting from the end, the first turn made by β that is different from α is tighter than the corresponding turn in α. More exactly, we require that $|\mathrm{LCS}(\alpha, \beta)| > 0$, and that

$$\angle(\alpha, |\alpha| - |\mathrm{LCS}(\alpha, \beta)|) < \angle(\beta, |\beta| - |\mathrm{LCS}(\alpha, \beta)|).$$

If $x = \alpha(|\alpha|) = \beta(|\beta|)$, we denote $\alpha \preceq_x \beta$ by a shorthand notation $\alpha \preceq \beta$.

Proof that $\preceq_x$ is a partial order in $\mathcal{L}_{sx}$.
Reflexivity. For all $\gamma \in \mathcal{L}_{sx}$, $\gamma \preceq_x \gamma$ directly by the first case of the definition.
Antisymmetry. Assume that $\alpha \preceq_x \beta \preceq_x \alpha$ for some $\alpha, \beta \in \mathcal{L}_{sx}$. Clearly the second and third case of the definition are strictly antisymmetric, i.e. if $\alpha \preceq_x \beta$ then $\beta \npreceq_x \alpha$. Thus $\alpha = \beta$.
Transitivity. Assume that $\alpha \preceq_x \beta \preceq_x \gamma$ for some $\alpha, \beta, \gamma \in \mathcal{L}_{sx}$. If $\alpha = \beta$ or $\beta = \gamma$, $\alpha \preceq_x \gamma$ is given directly by the assumption. If $|\alpha| < |\beta|$ or $|\beta| < |\gamma|$, we get that $|\alpha| < |\gamma|$ because $\alpha \preceq_x \beta$ implies $|\alpha| \leq |\beta|$ and $\beta \preceq_x \gamma$ implies $|\beta| \leq |\gamma|$. Thus the third case of the definition holds for both comparisons $\alpha \preceq_x \beta$ and $\beta \preceq_x \gamma$. Thus we know that $|\alpha| = |\beta| = |\gamma| =: L$. Let

$$x_1 = |\mathrm{LCS}(\alpha, \beta)|,$$
$$x_2 = |\mathrm{LCS}(\beta, \gamma)|.$$

By definition we know that $x_1, x_2 \in (0, L)$. There are three cases:

- If $x_1 = x_2$, then $\alpha_{x_1::} = \beta_{x_1::} = \gamma_{x_1::}$ and
$$\angle(\alpha, L - x_1) < \angle(\beta, L - x_1) < \angle(\gamma, L - x_1).$$
Furthermore, because $\angle(\alpha, L - x_1) \neq \angle(\gamma, L - x_1)$, for all $t > x_1$, $\alpha_{t::} \neq \gamma_{t::}$, and thus $|\mathrm{LCS}(\alpha, \gamma)| = x_1$. Thus $\alpha \preceq_x \gamma$.

- If $x_1 < x_2$, then $\alpha_{x_1::} = \beta_{x_1::} = \gamma_{x_1::}$. By definition we know that $\angle(\alpha, L - x_1) < \angle(\beta, L - x_1)$, and from the assumption $x_1 < x_2$ we know that $\angle(\beta, L - x_1) = \angle(\gamma, L - x_1)$. By the definition of LCS, for all $x_1 < t < x_2$ we know that $\alpha_{t::} \neq \beta_{t::} = \gamma_{t::}$, and thus we get that $|\mathrm{LCS}(\alpha, \gamma)| = x_1$. This combined with the earlier result that $\angle(\alpha, L - x_1) < \angle(\gamma, L - x_1)$ yields $\alpha \preceq_x \gamma$.

- The remaining case $x_1 > x_2$ is analogous to the previous case.

$\square$

Now using this refined length comparison $\preceq_x$, we are ready to define kth shortest paths.

Definition 5.3. Let $\gamma \in \mathcal{L}_{sx}$ for some $x \in P$. γ is a *kth shortest path* or *k-path* for some $k \in \mathbb{Z}_+$, if it has k non-strict predecessors and no incomparable elements in $\mathcal{L}_{sx}$, i.e.

$$|\{\alpha \in \mathcal{L}_{sx} \mid \alpha \preceq \gamma\}| = k,$$

and

$$\{\beta \in \mathcal{L}_{sx} \mid \beta \npreceq \gamma \text{ and } \gamma \npreceq \beta\} = \emptyset.$$

Denote the set of $x \in P$ such that there exists a k-path to x by P_k, and its complement $P \setminus P_k$ by T_k. Denote the unique k-path to $x \in P_k$ by $p_k(x)$.

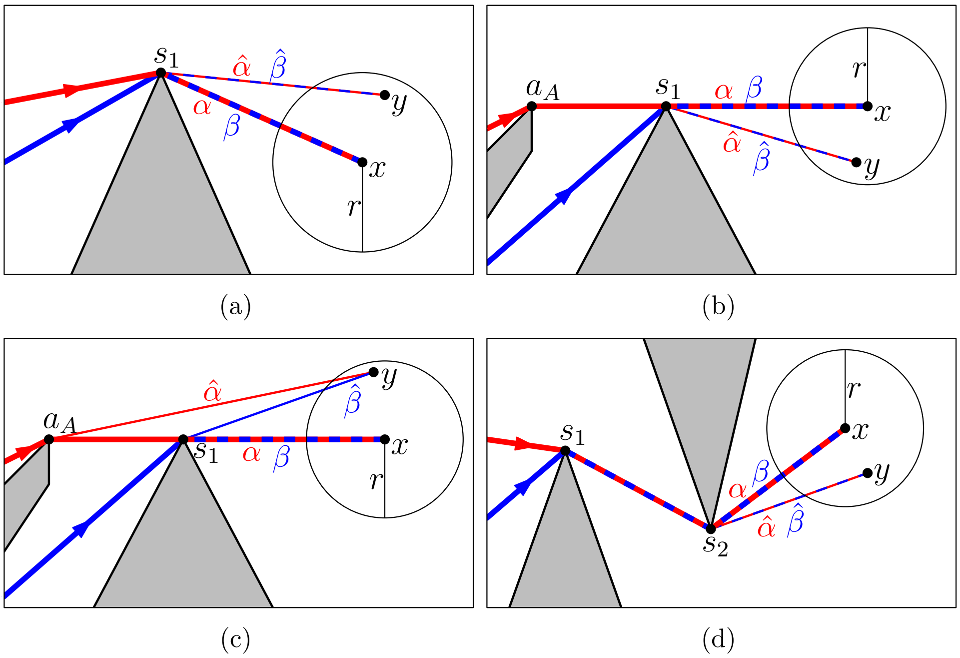


Figure 5.2: Examples of different cases of the proof of Lemma 5.4, where we assume that $\alpha, \beta \in \mathcal{L}_{sx}$ are distinct paths, $|\alpha| = |\beta|$ and $\alpha \preceq \beta$, and we want to prove that for some $r_x > 0$ it holds that if $y \in B_P(x, r_x)$, then by moving the endpoints of the paths to y (yielding paths $\hat{\alpha} = e_\alpha^{-1}(y)$ and $\hat{\beta} = e_\beta^{-1}(y)$), the ordering is preserved: $\hat{\alpha} \preceq \hat{\beta}$.

Proof that the k-path is unique. Let $k \in \mathbb{Z}_+$, $x \in P_k$ and $\alpha, \beta \in \mathcal{L}_{sx}$ and $\alpha \neq \beta$. Now if α and β are k-paths, then $\alpha \not\preceq \beta$ and $\beta \not\preceq \alpha$, because otherwise one of them would have more predecessors than the other. This is a contradiction with the definition that k-paths do not have incomparable elements. □

Now we will prove that for all k-paths, the endpoint can be moved by a small distance (keeping the path in $\mathcal{L}_s$), and the path is still a k-path, which means that we successfully achieved a consistent ordering of the paths. We will prove it with the help of the following two lemmas.

Lemma 5.4. *Let $x \in P$ and α and β be distinct paths of $\mathcal{L}_{sx}$ such that $\alpha \preceq \beta$ and $|\alpha| = |\beta|$. Use $r_x > 0$ and e_γ^{-1} from Theorem 3.13. Now for all $y \in B_P(x, r_x)$ it holds that $e_\alpha^{-1}(y) \preceq e_\beta^{-1}(y)$.*

Proof. Let $y \in B_P(x, r_x)$. Denote for all paths $\xi \in \mathcal{L}_{sx}$ the local inverse $e_\xi^{-1}(y)$ by $\hat{\xi}$. Let us write α and β as in Theorem 3.13 by listing all vertices the path touches, yielding

$$\alpha = [a_1, \ldots, a_A, s_1, \ldots, s_n, x]$$
$$\beta = [b_1, \ldots, b_B, s_1, \ldots, s_n, x],$$

where $[s_1, \ldots, s_n, x] = \mathrm{LCS}(\alpha, \beta)$. By the definition of $\preceq_x$ is we know that $\mathrm{LCS}(\alpha, \beta)$ is nontrivial, i.e. $A, B, n \geq 1$. Let t be the length of the part that is different in α and β, i.e.

$$\begin{aligned} t = |[a_1, \ldots, a_A, s_1]| \quad &= |[b_1, \ldots, b_B, s_1]| \\ = |\alpha| - |\mathrm{LCS}(\alpha, \beta)| &= |\beta| - |\mathrm{LCS}(\alpha, \beta)|. \end{aligned}$$

By the assumptions $\alpha \preceq \beta$ and $\alpha \neq \beta$ we know that $\angle(\alpha, t) < \angle(\beta, t)$. Let us prove the theorem in cases depending on the type of $\mathrm{LCS}(\alpha, \beta)$.

1. If $n = 1$: Because $\angle(\beta, t) > \angle(\alpha, t) \geq 180°$, the case 2 of Theorem 3.13 always holds if $\gamma = \beta$. Let us consider the cases of path α separately

   (a) If the case 2 or 3a of Theorem 3.13 holds for both $\gamma = \alpha$ and $\gamma = \beta$, then

   $$\hat{\alpha} = [a_1, \ldots, a_A, s_1, y]$$
   $$\hat{\beta} = [b_1, \ldots, b_B, s_1, y].$$

   Now we see that

   $$\angle(\hat{\beta}, t) - \angle(\hat{\alpha}, t) = \angle(\beta, t) - \angle(\alpha, t) > 0$$

   (see figures 5.2a and 5.2b). Thus $\hat{\alpha} \preceq \hat{\beta}$.

   (b) If the case 3b of Theorem 3.13 holds for $\gamma = \alpha$, then

   $$\hat{\alpha} = [a_1, \ldots, a_A, y]$$
   $$\hat{\beta} = [b_1, \ldots, b_B, s_1, y].$$

   Because the points a_A, s_1 and y are not collinear, we know that $|[a_A, s_1, y]| > |[a_A, y]|$. Therefore

   $$|\hat{\alpha}| < |[a_1, \ldots, a_A, s_1, y]| = |[b_1, \ldots, b_B, s_1, y]| = |\hat{\beta}|,$$

   which yields the claim $\hat{\alpha} \preceq \hat{\beta}$ (see Figure 5.2c).

2. If $n \geq 2$: In Theorem 3.13, in every case of the explicit representation of $e_\gamma(y)$ we notice that it depends only on the last three vertices of the path γ and will change or remove only the last two vertices, so there exists $s'_1, \ldots, s'_{n'}$ such that

   $$\hat{\alpha} = [a_1, \ldots, a_A, s'_1, \ldots, s'_{n'}]$$
   $$\hat{\beta} = [b_1, \ldots, b_B, s'_1, \ldots, s'_{n'}],$$

where $n' \geq 2$ and $s'_1 = s_1$. Now we see that

$$\angle(\hat{\beta}, t) - \angle(\hat{\alpha}, t) = \angle(\beta, t) - \angle(\alpha, t) > 0$$

(see Figure 5.2d). Thus $\hat{\alpha} \preceq \hat{\beta}$.

□

Lemma 5.5. *Let $x \in P$ and $L \geq 0$. Use $r_x > 0$ and e_γ^{-1} from Theorem 3.13. Now there exists $0 < r < r_x$ such that if $y \in B_P(x, r)$, then the map $\gamma \mapsto e_\gamma^{-1}(y)$ preserves the ordering of the lengths between paths of $\mathcal{L}_{sx}$ with length L and all the other paths, i.e. for all $\alpha, \beta \in \mathcal{L}_{sx}$ such that $|\alpha| = L$,*

- *$|\alpha| < |\beta|$ implies that $|e_\alpha^{-1}(y)| < |e_\beta^{-1}(y)|$,*
- *$|\alpha| > |\beta|$ implies that $|e_\alpha^{-1}(y)| > |e_\beta^{-1}(y)|$.*

Proof. Let D be the distance from L to the set of other path lengths in γ_{sx}. The closest length exists and $\delta > 0$ due to Theorem 3.10. Let us choose $r > 0$ such that $r < r_x$ and $r < D/2$. Let $y \in B_P(x, r)$ and $\alpha, \beta \in \mathcal{L}_{sx}$ such that $|\alpha| = L$ and $|\beta| \neq L$. Because of how we defined D, we know that $\Big||\beta| - |\alpha|\Big| \geq D > 2r$. By considering all the cases of the explicit formula for e_γ^{-1} in Theorem 3.13 it is easy to see that

$$\Big||e_\alpha^{-1}(y)| - |\alpha|\Big| \leq ||y - x|| < r,$$

$$\Big||e_\beta^{-1}(y)| - |\beta|\Big| \leq ||y - x|| < r.$$

Now we can bound $|e_\beta^{-1}(y)| - |e_\alpha^{-1}(y)|$ combining these inequalities and the result $\Big||\beta| - |\alpha|\Big| > 2r$ using the triangle inequality, yielding the claim that it has the same sign as $|\beta| - |\alpha|$. □

Theorem 5.6. *Let $k \in \mathbb{Z}_+$, $x \in P_k$ and $\gamma = p_k(x)$. Use $r_x > 0$ and e_γ^{-1} from Theorem 3.13. Now there exists a neighborhood $B_P(x, r)$ of x with $0 < r < r_x$ such that for all $y \in B_P(x, r)$, $e_\gamma^{-1}(y)$ is a k-path, i.e. $y \in P_k$ and $e_\gamma^{-1}(y) = p_k(y)$.*

Proof. We will use the $0 < r < r_x$ from Lemma 5.5 applied to x and $L = |\gamma|$. Let $y \in B_P(x, r)$.

Let $\alpha \in \mathcal{L}_{sx}$. If $|\alpha| \neq |\gamma|$, then we get from Lemma 5.5 that $|e_\alpha^{-1}(y)|$ and $|e_\gamma^{-1}(y)|$ are ordered the same as $|\alpha|$ and $|\gamma|$. If $|\alpha| = |\gamma|$, then because γ is a k-path, we know that α and γ are comparable by $\preceq$. Now Lemma 5.4 yields that the $\preceq$-ordering between $e_\alpha^{-1}(y)$ and $e_\beta^{-1}(y)$ is the same as that of α and β.

Thus because the map $\alpha \mapsto e_\alpha^{-1}(y)$ preserves the $\preceq$-ordering between the k-path γ and all the other paths, we get that $e_\gamma^{-1}(y)$ is also a k-path. □

As a simple corollary of the previous theorem, we also get that P_k is open in P. Note that generally, P_k is not open in $\mathbb{R}^2$, because it contains boundary points of P.

Corollary 5.7. *P_k is an open subset of P for all $k \in \mathbb{Z}_+$.*

5.3 Windows

The aim of this section was to develop a map from which one can directly read the kth shortest path from s to almost any $x \in P$. It turns out that the components of the set P_k almost form this kind of map: we only have to split some components by removing k-windows, which we will now define.

Definition 5.8. Let $k \in \mathbb{Z}_+$. Define the set of *k-windows* W_k as the set of $x \in P_k$ such that the last segment of the k-path $p_k(x)$ touches a vertex, i.e. there exists $v \in V$ such that $[v, x]$ is a suffix of $p_k(x)$ and $[v, x](t) \in V$ for some $t \in (0, |[v, x]|]$.

Define the kth shortest path map (or k-SPM in short) M_k of P by

$$M_k = P_k \setminus W_k.$$

See Figure 5.3 for an example illustrating what the sets M_k, T_k and W_k look like. The interactive k-SPM visualization applet [8] might also help in understanding the concepts. Now we are ready to prove that one can simply deduce the k-path $p_k(x)$ to x simply by finding the component of M_k it lies in.

Theorem 5.9. *Let $k \in \mathbb{Z}_+$. Let C be a path connected component of the k-SPM M_k. Then there exists $v \in V \cup \{s\}$ and $\gamma \in \mathcal{L}_{sv}$ such that for all $x \in C$,*

$$p_k(x) = \gamma[v, x].$$

Proof. Choose $x \in C$. Because $C \subset M_k \subset P_k$, the k-path $p_k(x)$ exists. Let $t \in [0, |\gamma|]$ the maximum such that $p_k(x)(t) \in V$ or $t = 0$. Denote $v = \gamma(t) \in V \cup \{s\}$ and $\gamma = p_k(x)_{[0,t]}$. Now $p_k(x) = \gamma[v, x]$. We will prove that for all $y \in C$, $p_k(y) = \gamma[v, y]$.

Choose an arbitrary curve within the component C from x to y, i.e. curve $f : [0, 1] \to C$ such that $f(0) = x$ and $f(1) = y$. Such a curve exists because C is a path-connected component. Define the set

$$X = \{t \in [0, 1] \mid p_k(f(t)) = \gamma[v, f(t)]\}.$$

We know that $0 \in X$ and it suffices to prove that $1 \in X$. Let $t = \sup X$. Let $\gamma = p_k(f(t))$. By Theorem 5.6, we know that there exists a r-neighborhood

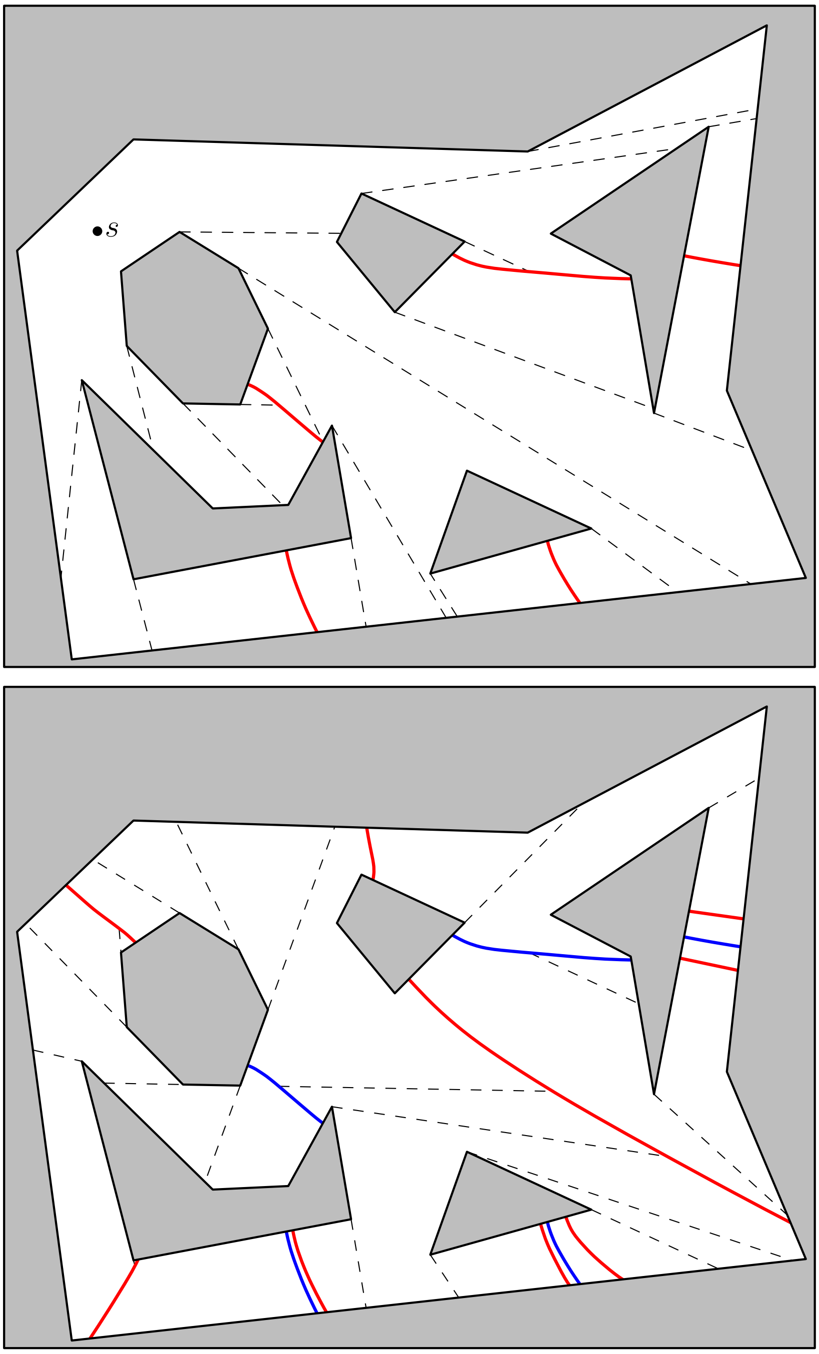


Figure 5.3: Illustration of kth shortest path map M_k for $k = 1$ and $k = 2$. The complement set T_k of P_k is drawn in red and blue ($T_{\leq k-1}$ in blue and $T_{\leq k}$ in red, see Definition 5.13). The set of k-windows W_k is drawn as dashed lines. It can be seen that W_k consists of non-intersecting line segments that are always contained within a single cell of P_k, as will be proven in Theorem 5.12. The full k-SPM M_k is the set P_k cut into parts by the k-windows.

of $f(t)$ such that for all $y \in B_P(f(t), r)$, $e_\gamma^{-1}(y) = p_k(y)$. Because $t = \sup X$ and f is continuous, for some $t' \in X$ we know that $f(t') \in B_P(f(t), r)$. Because $f(t) \in C$, $f(t)$ is not in the set of windows W_k and therefore case 1, 2 or 3a holds in Theorem 3.13. This means that since $p_k(f(t')) = \gamma[v, f(t')]$, also $p_k(f(t)) = \gamma[v, f(t)]$, which implies that $t \in X$. If $t < 1$, similarly we get that there exists $t' > t$ such that $f(t') \in B_P(f(t), r)$ and thus $p_k(f(t')) = \gamma[v, f(t')]$, which implies that $t' \in X$, a contradiction with the definition $t = \sup X$. Now we have proven that $t \in X$ and $t = 1$, and therefore $1 \in X$. □

Now we have established that the k-SPM M_k can be used to directly query kth shortest paths. However, apart from the intuition given by Figure 5.3 and the visualization applet [8], we still cannot assess the utility of it as a data structure, as we do not know how big a fraction of P is contained M_k, and what is the geometrical structure of M_k. We choose to study the geometrical structure of M_k through its complement $P \setminus M_k = T_k \cup W_k$.

First we prove that the set of k-windows W_k contributes at most $O(k|V|)$ new features to the complexity of the k-SPM by showing that they consist of $O(k|V|)$ non-intersecting line segments that subdivide the cells of P_k (as can be seen from the example Figure 5.3).

Lemma 5.10. *For all $k \in \mathbb{Z}_+$ and $v \in V$, there exists sets of size k,*

- $L_{v,k} \subset \{\gamma \in \mathcal{L}_{sv} \mid \gamma$ *comes to v from the left sector*$\}$
- $R_{v,k} \subset \{\gamma \in \mathcal{L}_{sv} \mid \gamma$ *comes to v from the right sector*$\}$

such that if $\gamma[v, x] \in \mathcal{L}_{sx}$ and $|\{\alpha \in \mathcal{L}_{sx} \mid \alpha \preceq \gamma[v, x]\}| \leq k$ (for example, if $\gamma[v, x]$ is a k-path), then:

- *If γ comes to v from the left sector, $\gamma \in L_{v,k}$.*
- *If γ comes to v from the right sector, $\gamma \in R_{v,k}$.*

Proof. We prove that a set $L_{v,k}$ such that the first claim holds exists. The other claim follows analogously. Let $(\gamma_1, \gamma_2, \ldots)$ be the sequence of paths of $\mathcal{L}_{sv}$ that come to v from the left sector ordered primarily by their lengths, then by the direction from which they come to v in counterclockwise order, and finally by $\preceq$. This ordering is total, because $\preceq$ always orders totally the paths that come to v from the same direction, since they have a nontrivial common suffix. We will prove that the claim holds if we choose

$$L_{v,k} = \{\gamma_1, \gamma_2, \ldots, \gamma_k\}.$$

Assume the contrary: $\gamma_i[v, x]$ has at most k predecessors in $\mathcal{L}_{sx}$ for some $i > k$. The paths $\gamma_1[v, x], \gamma_2[v, x], \ldots, \gamma_k[v, x]$ have distinct homotopy types by Theorem 3.17, because the paths $\gamma_1, \ldots, \gamma_k$ are distinct locally shortest paths.

Furthermore, also their shortcuts $\alpha_1, \ldots, \alpha_k$ have distinct homotopy types, as they have the same homotopy types as the originals $\gamma_1[v,x], \ldots, \gamma_k[v,x]$. Thus to get a contradiction, it suffices to prove that for all $j \in \{1, \ldots, k\}$ it holds that $\alpha_j \preceq \gamma_i[v,x]$, because then we have found k paths from s to x that are less than $\gamma_i[v,x]$ in terms of $\preceq$. By Theorem 3.18, we know that

$$|\alpha_j| \leq |\gamma_j[v,x]| \leq |\gamma_i[v,x]|.$$

Let us prove the contradiction in cases for all $j \in \{1, \ldots, k\}$:

- If $|\alpha_j| < |\gamma_i[v,x]|$, then directly by the definition of $\preceq$, $\alpha_j \preceq \gamma_i[v,x]$.
- If $|\alpha_j| = |\gamma_i[v,x]|$, then $\gamma_j[v,x]$ is a locally shortest path, because otherwise $\angle(\gamma_j[v,x], |\gamma_j|) < 180°$, and then clearly $|\alpha_j| < |\gamma_i[v,x]|$, because one can shorten $\gamma_i[v,x]$ simply by a small shortcut around v.

  Since $\gamma_j[v,x]$ is a locally shortest path, the shortcut is just the same path: $\alpha_j = \gamma_j[v,x]$. By the ordering of the paths $(\gamma_1, \gamma_2, \ldots)$, we know that
  $$\angle(\gamma_j[v,x], |\gamma_j|) \leq \angle(\gamma_i[v,x], |\gamma_i|).$$
  If equality holds in the inequality, then from the ordering of paths $(\gamma_1, \gamma_2, \ldots)$ we know that $\gamma_j \preceq \gamma_i$ and thus $\gamma_j[v,x] \preceq \gamma_i[v,x]$. Otherwise, we get that $\gamma_j[v,x] \preceq \gamma_i[v,x]$ directly from the definition of $\preceq$.

We proved the contradiction in both cases, which means that our choice of $L_{v,k}$ satisfies the claim. □

Lemma 5.11. *Let $\gamma \in \mathcal{L}_{sv}$ for some $v \in V$. Let $a, b \in P \setminus \{v\}$ such that $\alpha = \gamma[v,a]$ is a prefix of $\beta = \gamma[v,b]$ and $[v,b]$ does not contain any vertices apart from its endpoints.*

(a) If $\xi \in \mathcal{L}_{sa}$ such that $\xi \preceq \alpha$ and ϕ is the shortcut of $\xi[a,b]$, then $\phi \preceq \beta$.

(b) If $\xi \in \mathcal{L}_{sb}$ such that $\xi \succeq \beta$ and ϕ is the shortcut of $\xi[b,a]$, then $\phi \succeq \alpha$.

Proof.

(a) If $|\xi| = |\alpha|$, then because ξ and α are comparable, they have a common suffix $[v,a]$ and therefore ϕ and β have a common suffix $[v,b] = [v,a,b]$. This means that $\phi = \xi[a,b]$, and directly by the definition of $\preceq$, $\phi \preceq \alpha$. Otherwise, $|\xi| < |\alpha|$ and now by Theorem 3.18,

$$|\phi| \leq |\xi| + ||b-a|| < |\alpha| + ||b-a|| = |\beta|.$$

(b) If $|\xi| = |\beta|$, we get that $\phi \preceq \alpha$ analogously to the proof of part (a). Assume that $|\xi| > |\beta|$. If $|\phi| \leq |\alpha|$, then because ξ is the shortcut of $\phi[a, b]$, by Theorem 3.18 we get that

$$\begin{aligned}|\xi| &\leq |\phi| + ||b - a|| \\ &\leq |\alpha| + ||b - a|| = |\beta|,\end{aligned}$$

which is a contradiction with $|\xi| > |\beta|$. Therefore $|\phi| > |\alpha|$.

□

Theorem 5.12. *The set of k-windows W_k consists of at most $2k|V|$ non-intersecting line segments for all $k \in \mathbb{Z}_+$. The endpoints of the segments are contained in $T_k \cup \partial P$.*

Proof. Let $\gamma \in \mathcal{L}_{sv}$ and $v \in V$. Define the γ-window denoted by W_γ as the set of points $x \in P$ such that $\gamma[v, x]$ is a path contained in P and $v = x$ or $\angle(\gamma[v, x], |\gamma|) = 180°$, i.e. $[v, x]$ is an extension to the last segment of the path γ. Therefore, W_γ is the line segment obtained by ray-shooting from v towards the direction of the end of γ. By the definition of the set W_k of k-windows, for all $k \in \mathbb{Z}_+$

$$W_k = \bigcup_{v \in V} \bigcup_{\gamma \in \mathcal{L}_{sv}} W_{k,\gamma},$$

where $W_{k,\gamma}$ is the set of of points of W_γ such that γ extended to that point is a k-path, i.e.

$$W_{k,\gamma} = \{\gamma[v, x] = p_k(x) \mid x \in P_k \cap W_\gamma\}.$$

By Lemma 5.10, if $\gamma \notin L_{v,k} \cup R_{v,k}$, then $W_{k,\gamma} = \emptyset$, and thus we can write W_k as a smaller union

$$W_k = \bigcup_{v \in V} \bigcup_{\gamma \in (L_{v,k} \cup R_{v,k})} W_{k,\gamma}.$$

Now $|L_{v,k} \cup R_{v,k}| \leq 2k$ and we know that if $a, b \in V$ and $\alpha \in \mathcal{L}_{sa}, \beta \in \mathcal{L}_{sb}$ are distinct paths, then $W_{k,\alpha} \cap W_{k,\beta} = \emptyset$, because there cannot be multiple k-paths to the same point. We have now proven that the set W_k consists of at most $2k|V|$ non-intersecting subsets of line segments, and the only part missing from the first claim is that the sets $W_{k,\gamma}$ are actually connected.

Thus to prove the first claim of the theorem, it suffices to prove that if $v \in V$ and $\gamma \in \mathcal{L}_{sv}$, then for all $x, y \in W_{k,\gamma}$ and $t \in [0, 1]$, $[x, y]'(t) \in W_{k,\gamma}$. Without loss of generality, we may assume that x is closer to v than y. Denote $m(t) = [x, y]'(t)$ and $\gamma_t = \gamma[v, m(t)]$. From the assumption we know that γ_0 and γ_1 are k-paths, and we want to prove that γ_t is a k-path for all $0 < t < 1$.

Define for all $a, b \in [0, 1]$ a mapping $f_{ab} : \mathcal{L}_{sm(a)} \to \mathcal{L}_{sm(b)}$ such that $f_{ab}(\xi)$ is the shortcut of $\xi[m(a), m(b)]$ (as defined in Theorem 3.17). Now for all $a, b, c \in [0, 1]$, $f_{bc}(f_{ab}(\xi))$ is the shortcut of $\xi[m(a), m(b), m(c)]$, which is equal to the shortcut of $\xi[m(a), m(c)]$ for all $\xi \in \mathcal{L}_{sm(a)}$. Therefore $f_{ac} = f_{bc} \circ f_{ab}$. As a special case, by setting $c = a$, we get that f_{ba} is the inverse of f_{ab}, which yields that f_{ab} is bijective for all $a, b \in [0, 1]$. We get that $f_{ab}(\gamma_a) = \gamma_b$ in the case $a \leq b$ directly from the definition and in the case $a > b$ using the fact that f_{ba} is the inverse f_{ab}.

Let $0 \leq a \leq b \leq 1$. We get the following rules about how the functions f_{ab} and f_{ba} preserve the order of paths as direct consequences of Lemma 5.11:

(a) If $\xi \in \mathcal{L}_{sm(a)}$ such that $\xi \preceq \gamma_a$, then $f_{ab}(\xi) \preceq \gamma_b$.

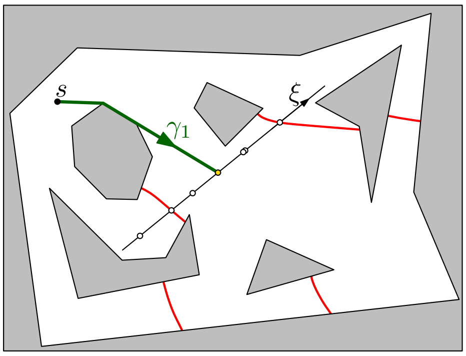
(b) If $\xi \in \mathcal{L}_{sm(b)}$ such that $\xi \succeq \gamma_b$, then $f_{ba}(\xi) \succeq \gamma_a$.

Now using these rules, we can prove the claim: Let $\xi \in \mathcal{L}_{sx} \setminus \{\gamma_0\}$. If $\xi \preceq \gamma_0$, then by rule (a), we know that $f_{0t}(\xi) \preceq \gamma_t$. Assume the other case: $\xi \succeq \gamma_0$. By rule (a), we get that f_{01} maps the set of paths of $\mathcal{L}_{sx}$ smaller than γ_0 to the set of paths of $\mathcal{L}_{sy}$ smaller than γ_1. Since γ_0 and γ_1 are k-paths, those sets have equal sizes and by the bijectivity of f_{01} this means that f_{01} preserves the $\preceq$-ordering between γ_0 and all the other paths of $\mathcal{L}_{sx}$. Therefore $f_{01}(\xi) \succeq \gamma_1$. Now using rule (b), we know that $f_{1t}(f_{01}(\xi)) \succeq f_{1t}(\gamma_1)$, which in simplified form means that $f_{0t}(\xi) \succeq \gamma_t$. Combining the results, we know that f_{0t} preserves the ordering between γ_0 and all the other paths of $\mathcal{L}_{sx}$, and therefore because γ_0 is a k-path, γ_t is also a k-path.

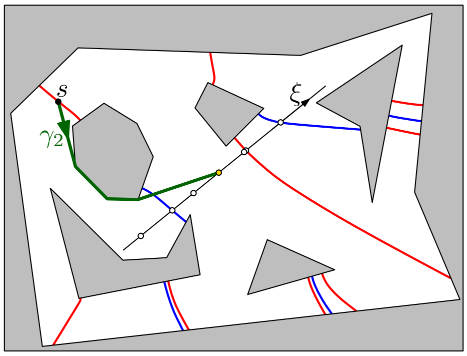
Now let us prove that the endpoints of the segments are contained in $T_k \cup \partial P$. Assume the contrary: for some $v \in V$ and $\gamma \in \mathcal{L}_{sv}$, an endpoint of $W_{k,\gamma}$ is not in $T_k \cup \partial P$. Let $p, q \in \mathbb{R}$ such that $\xi(p, q) \subset W_{k,\gamma} \subset \xi[p, q]$ where ξ is the path traversing the segment W_γ. Now $\xi(q) \notin T_k \cup \partial P$ or $\xi(p) \notin T_k \cup \partial P$. Let us assume the former case, because the latter case can be handled analogously. Now because $\xi(q) \notin \partial P$, we know that $q < |\xi|$. Because $\xi(p) \notin T_k$, $\xi(p) \in P_k$ and thus there exists $\alpha = p_k(\xi(q))$. By Theorem 5.6 we now see that because $\xi(q)$ is a limit point of $W_{k,\gamma}$, it holds that $\alpha = \gamma[v, \xi(q)]$ and furthermore that there exists $t > q$ such that $p_k(\xi(t)) = e_\alpha^{-1}(\xi(t)) = \gamma[v, \xi(t)]$. Thus $\xi(t) \in W_{k,\gamma}$ which is a contradiction with the assumption $W_{k,\gamma} \subset \xi[p, q]$. □

5.4 Walls

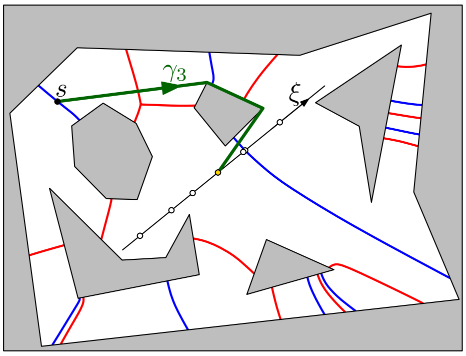
Intuitively, as we move a point x around P, the lengths of the paths in $\mathcal{L}_{sx}$ change continuously, and the k-path can change discontinuously for two reasons: if the k-path and the $(k-1)$-path switch places, or if the k-path and the $(k+1)$-path switch places. The first case happens exactly when the set of $k - 1$ shortest paths changes, and the second case exactly when



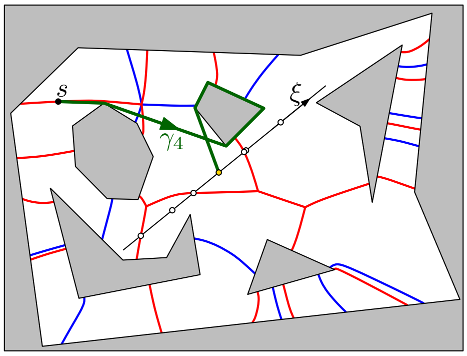
(a) P_1



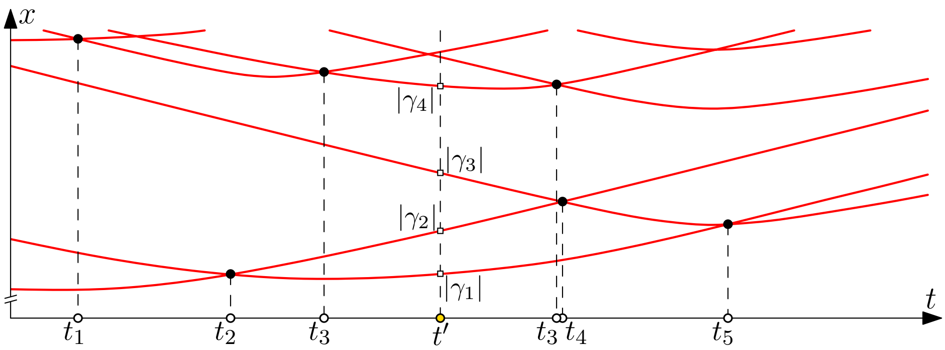
(b) P_2



(c) P_3



(d) P_4



(e) Points (t, x) such that $x \in \{|\gamma| \mid \gamma \in \mathcal{L}_{s\xi(t)}\}$.

Figure 5.4: Consider the lengths of the locally shortest paths from s to points on a segment ξ. The lengths of all the paths in $\mathcal{L}_{s\xi(t)}$ as $\xi(t)$ traverses the segment are drawn in (e). Here $t_1, \ldots, t_5$ are the points in which the four shortest paths in $\mathcal{L}_{s\xi(t)}$ change order. In (a)-(d) we draw P_k for $k \in \{1, 2, 3, 4\}$ such that the complement T_k is drawn in red and blue: red in points where the k-path and the $(k+1)$-path switch places ($T_{\leq k-1}$ in Definition 5.13), and blue in points where the k-path and the $(k-1)$-path switch places ($T_{\leq k}$ in Definition 5.13). Note how $t_1, \ldots, t_5$ correspond to the intersections between ξ and $T_{\leq 1}, \ldots, T_{\leq 4}$. We also draw the k-path $\gamma_k = p_k(\xi(t'))$ to P_k in (a)-(d).

the set of k shortest paths changes. See Figure 5.4 for an example and the visualization applet [8] for more intuition. Because of this connection, we will use this section to investigate the geometrical structure of the set of points in which the set of k shortest paths changes.

Definition 5.13. Let $\Gamma \subset \mathcal{L}_{sx}$ for some $x \in P$. We define that Γ is a *set of kth shortest paths* or *k-pathset* to x, if $|\Gamma| = k$ and all other paths of $\mathcal{L}_{sx}$ are larger in terms of $\preceq$ than the paths of Γ, i.e. for all $\gamma \in \Gamma$ and $\alpha \in \mathcal{L}_{sx} \setminus \Gamma$ it holds that $\gamma \preceq \alpha$. Denote the set of $x \in P$ such that there exists a k-pathset to x by $P_{\leq k}$, and its complement $P \setminus P_{\leq k}$ by $T_{\leq k}$. We say that $T_{\leq k}$ the set of *k-walls*. Denote the unique k-pathset to $x \in P_{\leq k}$ by $p_{\leq k}(x)$.

Proof that the k-pathset is unique. This follows directly from the following lemma. □

Lemma 5.14. *Let A and B be pathsets to $x \in P$. Now $A \subset B$ or $B \subset A$. Consequently, if $1 \leq a \leq b$ and $x \in P_{\leq a} \cap P_{\leq b}$, then $p_{\leq a}(x) \subset p_{\leq b}(x)$.*

Proof. If the claim does not hold, there exists $\alpha \in A$ and $\beta \in B$ such that $\alpha \notin B$ and $\beta \notin A$. By the definition of k-pathsets we now know that $\alpha \preceq \beta$ and $\beta \preceq \alpha$, which implies that $\alpha = \beta$. This is a contradiction with the fact that $\alpha \in A$ and $\beta \notin A$. □

See Figure 5.3 for an example of the set of k-walls (drawn in red and blue). Now we can prove that the intuitive relation between paths $\{p_k\}$ and pathsets $\{p_{\leq k}\}$ holds with our definitions.

Theorem 5.15. *For all $k \in \mathbb{Z}_+$,*

$$P_k = P_{\leq k} \cap P_{\leq k-1},$$

or equivalently with complements,

$$T_k = T_{\leq k} \cup T_{\leq k-1},$$

For all $x \in P_k$,

$$p_{\leq k}(x) = p_{\leq k-1}(x) \cup \{p_k(x)\}.$$

Proof. Let $k \in \mathbb{Z}_+$ and $x \in P_k$. Let $A = \{\alpha \in \mathcal{L}_{sx} \mid \alpha \preceq p_k(x)\}$ and $B = A \setminus \{p_k(x)\}$. Now by definition, $|A| = k$ and therefore $|B| = k - 1$. Also by definition, there are no elements in $\mathcal{L}_{sx}$ that are incomparable with $p_k(x)$, which means that for all $\alpha \in \mathcal{L}_{sx} \setminus A$, $\alpha \succeq p_k(x)$. Because $p_k(x) \preceq p_k(x)$, this holds also for all $\alpha \in \mathcal{L}_{sx} \setminus B$. Therefore A is a k-pathset and B is a $(k-1)$-pathset. Thus $x \in P_{\leq k} \cap P_{\leq k-1}$ and $p_{\leq k}(x) = p_{\leq k-1}(x) \cup \{p_k(x)\}$. Now the only claim that remains to be proven is that if $x \in P_{\leq k} \cap P_{\leq k-1}$, then $x \in P_k$. By Lemma 5.14 we know that $p_{\leq k-1}(x) \subset p_{\leq k}(x)$. Let $\gamma \in \mathcal{L}_{sx}$ be the path such that $p_{\leq k}(x) = p_{\leq k-1}(x) \cup \{\gamma\}$. Now because $p_{\leq k}(x)$ is a pathset, for all $\alpha \in \mathcal{L}_{sx} \setminus p_{\leq k}(x)$ it holds that $\gamma \preceq \alpha$, which implies that $\alpha \npreceq \gamma$. Because $p_{\leq k-1}$ is a pathset, for all $\alpha \in p_{\leq k-1}$ it holds that $\alpha \preceq \gamma$. Combining these results, we get that the set of predecessors of γ is the set $p_{\leq k}$ which has size k, and no element of $\mathcal{L}_{sx}$ is incomparable with γ. By the definition of k-paths this means that γ is a k-path and thus $x \in P_k$. □

Now, to analyze the complexity and the structure of T_k, let us explore what the set of k-walls $T_{\leq k}$ consists of.

Definition 5.16. Let $x \in P$ and $k \in \mathbb{Z}_+$.

- Denote the kth element of the sorted list of the lengths of all the paths in $\mathcal{L}_{sx}$ (possibly containing duplicates) by $\mathcal{L}_{sx}(k)$.
- Denote the set paths in $\mathcal{L}_{sx}$ with lengths in set X by $\mathcal{L}_{sx,X}$, i.e.

$$\mathcal{L}_{sx,X} = \{\gamma \in \mathcal{L}_{sx} \mid |\gamma| \in X\}.$$

Theorem 5.17. *Let $x \in P$, $k \in \mathbb{Z}_+$ and $L = \mathcal{L}_{sx}(k)$. Now $x \in T_{\leq k}$ if and only if $|\mathcal{L}_{sx,[0,L]}| > k$ and there exists paths $\alpha, \beta \in \mathcal{L}_{sx,\{L\}}$ going to x from different directions (i.e. $\alpha \neq \beta$ and $|\mathrm{LCS}(\alpha, \beta)| = 0$).*

Proof. By the definition of L we know that $|\mathcal{L}_{sx,[0,L]}| \geq k$. If $|\mathcal{L}_{sx,[0,L]}| = k$, then the set $\mathcal{L}_{sx,[0,L]}$ is a k-pathset. If all the paths of $\mathcal{L}_{sx,\{L\}}$ share a common suffix of nonzero length, then by Definition 5.2 they are totally ordered by $\preceq$, and by ordering them to sequence $(\gamma_1, \gamma_2, \dots, \gamma_{|\mathcal{L}_{sx,\{L\}}|})$ we get that the set $\mathcal{L}_{sx,[0,L)} \cup \{\gamma_1, \dots, \gamma_{k-|\mathcal{L}_{sx,[0,L)}|}\}$ is a k-pathset. Both cases imply that $x \notin T_{\leq k}$.

We still need to prove that if $|\mathcal{L}_{sx,[0,L]}| > k$ and there exists $\alpha, \beta \in \mathcal{L}_{sx,\{L\}}$ such that $|\mathrm{LCS}(\alpha, \beta)| = 0$, then $x \in T_{\leq k}$. Assume the contrary: $\Gamma \subset \mathcal{L}_{sx}$ is a k-pathset. We know that $\Gamma \subset \mathcal{L}_{sx,[0,L]}$, because if for some $\gamma \in \Gamma$ it holds that $|\gamma| > L$, then we know that the set of shorter paths $\mathcal{L}_{sx,[0,L]}$ is also a subset of Γ, which is a contradiction, because then $|\Gamma| > k$. Let $A = \Gamma \cap \mathcal{L}_{sx,\{L\}}$ and $B = \mathcal{L}_{sx,\{L\}} \setminus \Gamma$. Because we assumed that $|\mathcal{L}_{sx,[0,L]}| > k$ and by definition $|\mathcal{L}_{sx,[0,L)}| < k$, it holds that A and B are nonempty sets. Because Γ is a pathset, for all $\alpha \in A$ and $\beta \in B$ it holds that $\alpha \preceq \beta$ and thus by Definition 5.2, $|\mathrm{LCS}(\alpha, \beta)| > 0$, which now yields the contradiction that all the paths of $A \cup B = \mathcal{L}_{sx,\{L\}}$ share a common suffix of nonzero length. □

In the above theorem, $x \in T_{\leq k}$ if there exists two paths $\xi, \phi \in \{\gamma_a, \dots, \gamma_b\}$ such that $|\mathrm{LCS}(\xi, \phi)| = 0$. Now if we remove the last segments of ξ and ϕ, we get that $\xi = \alpha[a, x]$ and $\phi = \beta[b, x]$ where $a, b \in V \cup \{s\}$ are distinct, $\alpha \in \mathcal{L}_{sa}$ and $\beta \in \mathcal{L}_{sb}$. This means that x is on the curve of points y for which

$$||y - a|| + |\alpha| = ||y - b|| + |\beta|.$$

As we will be forming the set $T_{\leq k}$ from these curves, let us name them.

Definition 5.18. Let $a, b \in V$ such that $a \neq b$ and $A, B \geq 0$. Now define the bisector between (a, A) and (b, B) by

$$\mathcal{B}(a, A; b, B) = \{y \in \mathbb{R}^2 \mid ||y - a|| + A = ||y - b|| + B\}.$$

It turns out that these bisector curves are always branches of a hyperbola, because the geometric definition of hyperbolas is that they are the curves defined as a set of points such that the absolute value of the difference of distances to two given points is a given constant. However, we only need the fact that they are well-behaved infinite curves.

Theorem 5.19. *Let $a, b \in V$ such that $a \neq b$, $A, B \geq 0$ and $|B - A| < ||b - a||$, then $\mathcal{B}(a, A; b, B)$ is an infinite curve, i.e. there exists a continuous function $f : \mathbb{R} \to \mathbb{R}^2$ such that*

$$f\mathbb{R} = \mathcal{B}(a, A; b, B).$$

Denote this f uniquely by $f_{\mathcal{B}(a,A;b,B)}$. $\mathcal{B}(a, A; b, B)$ is a very well-behaved curve: it is a subset of a hyperbola, and the function f satisfies the following properties:

- *f is injective.*
- *$|f(t)| \to \infty$ as $|t| \to \infty$.*
- *f is continuously differentiable.*

Proof. The fact that $\mathcal{B}(a, A; b, B)$ is a subset of a hyperbola follows directly from the definition. Let us now construct the function f. Consider the case $a = (-1, 0)$, $b = (1, 0)$ and $0 \leq B - A < 2$. The other cases follow by symmetry, scaling and rotation. Now $(x, y) \in \mathcal{B}(a, A; b, B)$ if

$$||(x, y) - a|| - ||(x, y) - b|| = B - A.$$

Note that if $x < 0$, the equation cannot hold, because the left hand side is negative and the right hand side is nonnegative. Assume that $x \geq 0$. Now both sides are nonnegative. Let us expand the left-hand side and square both sides:

$$\left(\sqrt{(x+1)^2 + y^2} - \sqrt{(x-1)^2 + y^2}\right)^2 = (B - A)^2.$$

yielding an equivalent equation

$$(x+1)^2 + (x-1)^2 + 2y^2 - 2\sqrt{((x+1)^2 + y^2)((x-1)^2 + y^2)} = (B - A)^2.$$

Expanding and reorganizing, we get that equivalently

$$2\sqrt{(x^2 + y^2 + 1)^2 - 4x^2} = 2(x^2 + y^2 + 1) - (B - A)^2.$$

If $2(x^2+y^2+1) \geq (B-A)^2$, both sides are nonnegative, and we get an equivalent equation by squaring both sides:

$$4(x^2+y^2+1)^2 - 16x^2 = 4(x^2+y^2+1)^2 + (B-A)^4 - 4(x^2+y^2+1)(B-A)^2.$$

This simplifies to

$$(16-4(B-A)^2)x^2 = 4(y^2+1)(B-A)^2 - (B-A)^4.$$

Because of the assumption $0 \leq B-A < 2$, the factor $16-4(B-A)^2$ is positive and the right-hand side is nonnegative. Because $x \geq 0$, we can solve x for a fixed y

$$x = (B-A)\sqrt{\frac{4(y^2+1)-(B-A)^2}{16-4(B-A)^2}}. \tag{4}$$

By a substitution we see that for any fixed y, if x is given by the above equation, $2(x^2+y^2+1) \geq (B-A)^2$:

$$\begin{aligned}
2(x^2+y^2+1) &= 2\left((B-A)^2\frac{4(y^2+1)-(B-A)^2}{16-4(B-A)^2} + y^2+1\right) \\
&\leq 2\left((B-A)^2\frac{4(0+1)-(B-A)^2}{16-4(B-A)^2} + 0+1\right) \\
&\leq 2\left((B-A)^2/4+1\right) = (B-A)^2/2+2 \geq (B-A)^2.
\end{aligned}$$

Therefore (4) is the only solution for the original equation $||(x,y)-a|| - ||(x,y)-b|| = B-A$ for fixed y. Now if we define f by

$$f(t) = \left(\frac{B-A}{2}\sqrt{\frac{4(t^2+1)-(B-A)^2}{4-(B-A)^2}}, t\right),$$

then clearly f satisfies the conditions. $\square$

Now let us define the k-bisectors, the building blocks of the set of k-walls $T_{\leq k}$, as the parts of the bisectors that contribute to the set $T_{\leq k}$.

Definition 5.20. Denote the set of pairs $(a,A) \in V \times [0,\infty)$ such that $\mathcal{L}_{sa,\{A\}} \neq \emptyset$ by $\mathcal{L}_{s\cdot,\{\cdot\}}$. Let $k \in \mathbb{Z}_+$ and $(a,A),(b,B) \in \mathcal{L}_{s\cdot,\{\cdot\}}$ such that $|B-A| < ||b-a||$. Define the *k-bisector between* (a,A) *and* (b,B) denoted by $T_{\leq k}(a,A;b,B)$ such that $x \in T_{\leq k}(a,A;b,B)$ if the following conditions hold:

- $x \in \mathcal{B}(a,A;b,B)$, i.e. there exists

$$L = A + ||x-a|| = B + ||x-b||.$$

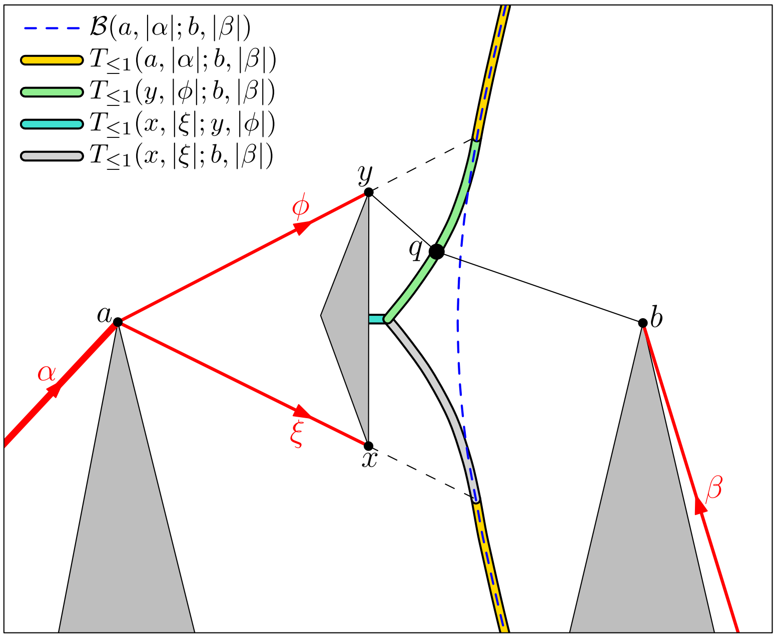


Figure 5.5: In the example configuration of the figure, α, β, ξ and ϕ are the shortest paths to vertices a, b, x and y, respectively. Furthermore, the shortest paths to x and y comes directly from a, and thus $\xi = \alpha[a,x]$ and $\phi = \alpha[a,y]$. In this case, we get four 1-bisectors, drawn in different colors. For example, the bisector $T_{\leq 1}(y,|\phi|;b,|\beta|)$ consists of the points $q \in \mathcal{B}(y,|\phi|;b,|\beta|)$ such that $\phi[y,q]$ and $\beta[b,q]$ are locally shortest paths, and $|\mathcal{L}_{sq,[0,L_q)}| < 1 < |\mathcal{L}_{sq,[0,L_q]}|$, where $L_q = |\phi| + ||q-y|| = |\beta| + ||q-b||$. The 1-bisector $T_{\leq 1}(a,|\alpha|;b,|\beta|)$ is disconnected, because the middle obstacle blocks visibility from a to some of the points of $\mathcal{B}(a,|\alpha|;b,|\beta|)$.

- $|\mathcal{L}_{sx,[0,L)}| < k < |\mathcal{L}_{sx,[0,L]}|$.
- There exists $\alpha \in \mathcal{L}_{sa,\{A\}}$ and $\beta \in \mathcal{L}_{sb,\{B\}}$ such that $\alpha[a,x], \beta[b,x] \in \mathcal{L}_{sx}$.

Note that this definition is symmetric with respect to the ordering of (a,A) and (b,B), i.e. $T_{\leq k}(a,A;b,B) = T_{\leq k}(b,B;a,A)$. For convenience, denote the set of unordered pairs $\{(a,A),(b,B)\} \subset \mathcal{L}_{s\cdot,\{\cdot\}}$ such that $|B-A| < ||b-a||$ by $T_{\leq k}(\cdot,\cdot)$. See Figure 5.5 for an example configuration of 1-bisectors.

Theorem 5.21. *The union of the k-bisectors defined in Definition 5.20 gives the whole set of k-walls $T_{\leq k}$, i.e.*

$$T_{\leq k} = \bigcup \{T_{\leq k}(a,A;b,B) \mid \{(a,A),(b,B)\} \in T_{\leq k}(\cdot,\cdot)\}.$$

Proof. Assume that $x \in T_{\leq k}$. We set $L = \mathcal{L}_{sx}(k)$, which immediately yields that $|\mathcal{L}_{sx,[0,L)}| < k$. By Theorem 5.17 we know that also $|\mathcal{L}_{sx,[0,L]}| > k$ and there exists $\xi, \phi \in \mathcal{L}_{sx,\{L\}}$ such that $|\text{LCS}(\xi, \phi)| = 0$. Because $|\xi| = |\phi|$, by Theorem 2.12 we know that $\alpha \neq [s, x] \neq \beta$. Therefore there exists $a, b \in V \setminus \{x\}$, $\alpha \in \mathcal{L}_{sa}$ and $\beta \in \mathcal{L}_{sb}$ such that $\xi = \alpha[a, x]$ and $\phi = \beta[b, x]$. Denote $A = |\alpha| = L - ||a - x||$ and $B = |\beta| = L - ||b - x||$. Now $(a, A), (b, B) \in \mathcal{L}_{s\cdot,\{\cdot\}}$. Now by the triangle inequality,

$$|B - A| = |||a - x|| - ||b - x||| \leq ||(a - x) - (b - x)|| = ||b - a||,$$

Furthermore, $|B - A| = ||b - a||$ cannot hold because $|\text{LCS}(\xi, \phi)| = 0$. Thus $|B - A| < ||b - a||$. Now we have found that all the conditions of Definition 5.20 hold, i.e. $x \in T_{\leq k}(a, A; b, B)$.

For the inverse direction, assume that for some $(a, A), (b, B) \in \mathcal{L}_{s\cdot,\{\cdot\}}$ it holds that $|B - A| < ||b - a||$ and $x \in T_{\leq k}(a, A; b, B)$. Now by Definition 5.20 we know that if we set $L = ||x - a|| + A = ||x - b|| + B$, then $\mathcal{L}_{sx,[0,L)} < k$ and $\mathcal{L}_{sx,[0,L]} > k$, which implies that $L = \mathcal{L}_{sx}(k)$. Furthermore, we can choose $\alpha \in \mathcal{L}_{sa,\{A\}}$ and $\beta \in \mathcal{L}_{sb,\{B\}}$ such that $\alpha[a, x], \beta[b, x] \in \mathcal{L}_{sx,\{L\}}$. To prove that $x \in T_{\leq k}$ using Theorem 5.17, we only need to prove that $|\text{LCS}(\alpha[a, x], \beta[b, x])| = 0$.

By the symmetry of (a, A) and (b, B) we may assume that $A \leq B$, which yields that $||x - a|| \geq ||x - b||$. If $|\text{LCS}(\alpha[a, x], \beta[b, x])| > 0$, then a and b are in the same direction from x, which yields that $||x - b|| + ||b - a|| = ||x - a||$. By plugging in the identities $||x - a|| = L - A$ and $||x - b|| = L - B$ we get that $||b - a|| = B - A = |B - A|$, which is a contradiction with the condition $|B - A| < ||b - a||$. Thus $\text{LCS}(\alpha[a, x], \beta[b, x]) = 0$. □

We have now successfully decomposed the complicated set $T_{\leq k}$ into the sets $T_{\leq k}(a, A; b, B)$ that are possibly disconnected subcurves of hyperbolic bisector curves. However, they still are a possibly large number of possibly intersecting curves. Let us investigate the set of possible intersection points and endpoints of the components of the sets $T_{\leq k}(a, A; b, B)$.

Definition 5.22. Let $k \in \mathbb{Z}_+$ and $x \in P$. Now we define that x is a *k-vertex* if it is an intersection point of either two k-bisectors or a k-bisector and the boundary of P, i.e. one of the following holds:

- There exists $\{(a, A), (b, B)\} \in T_{\leq k}(\cdot, \cdot)$ such that $x \in T_{\leq k}(a, A; b, B)$ and $x \in \partial P$.
- There exists distinct $\{(a, A), (b, B)\}, \{(c, C), (d, D)\} \in T_{\leq k}(\cdot, \cdot)$ such that
$$x \in T_{\leq k}(a, A; b, B) \cap T_{\leq k}(c, C; d, D).$$

We will prove that the k-bisectors consist of closed hyperbolic segments, and all of their endpoints are k-vertices. From the definition of k-vertices it is already clear that these segments can only intersect in k-vertices. This structure will be useful in Subsection 5.5, because if we form $T_{\leq k}$ (and ultimately, the boundary of $M_k = P \setminus (T_{\leq k} \cup T_{\leq k-1} \cup W_k)$) from a finite network of vertices and edges, we can query kth shortest paths to point x by finding out which cell of M_k it lies in using standard point location query algorithms.

Theorem 5.23. *Let $\{(a, A), (b, B)\} \in T_{\leq k}(\cdot, \cdot)$. Now if $f = f_{\mathcal{B}(a,A;b,B)}$ is as defined in Theorem 5.19, then the set*

$$I = f^{-1}T_{\leq k}(a, A; b, B)$$

is compact.

Proof. Because $|f(t)| \to \infty$ when $|t| \to \infty$ and P is bounded, $I \subset f^{-1}P$ is bounded. Thus it suffices to prove that for all the boundary points $t \in \partial I$ it holds that $t \in I$, i.e. $f(t) \in T_{\leq k}(a, A; b, B)$.

Because $f\mathbb{R} = \mathcal{B}(a, A; b, B)$, $A + ||f(t) - a|| = B + ||f(t) - b||$ for all $t \in \mathbb{R}$. Denote this number by $L(t)$. Now clearly L is a continuous function $\mathbb{R} \to \mathbb{R}$.

Now if $t \in \partial I$, there exists a sequence $(t_i)_{i=1}^{\infty}$ such that $t_i \in I$ for all $i \in \mathbb{Z}_+$ and $t_i \to t$ when $i \to \infty$. Because $t_i \in I$, there exists $\alpha_i \in \mathcal{L}_{sa,\{A\}}$ such that $\alpha_i[a, f(t_i)] \in \mathcal{L}_{sf(t_i),\{L(t_i)\}}$ for all $i \in \mathbb{Z}_+$. Theorem 3.10 yields that $\mathcal{L}_{sa,\{A\}}$ is finite, and thus some $\alpha \in \mathcal{L}_{sa,\{A\}}$ appears infinitely many times in the sequence $(\alpha_i)_{i=1}^{\infty}$. Thus by only taking the subsequence of $i \in \mathbb{Z}_+$ where $\alpha_i = \alpha$, we can choose $(t_i)_{i=1}^{\infty}$ such that for all $i \in \mathbb{Z}_+$, $\alpha[a, f(t_i)] \in \mathcal{L}_{sf(t_i),L(t_i)}$. Because P is closed and we get the path $\alpha[a, f(t)]$ as a limit of the paths $\alpha[a, f(t_i)]$ when $i \to \infty$, it is easy to see that $\alpha[a, f(t)] \in \mathcal{L}_{sf(t),\{L(t)\}}$. Similarly we find that there exists $\beta \in \mathcal{L}_{sb,\{B\}}$ such that $\beta[b, f(t)] \in \mathcal{L}_{sf(t),\{L(t)\}}$.

To finish the proof that $f(t) \in T_{\leq k}(a, A; b, B)$ we still need to prove the other condition: that $|\mathcal{L}_{sf(t),[0,L(t))}| < k$ and $|\mathcal{L}_{sf(t),[0,L(t)]}| > k$. Because $t \in \partial I$ and f is continuous, there exists a sequence $(t_i)_{i=1}^{\infty}$ such that for all $i \in \mathbb{Z}_+$, $t_i \in I$ and $||f(t_i) - f(t)|| < 1/i$. By Lemma 3.7 we may choose the elements of the sequence to be close enough to t such that $[f(t_i), f(t)]$ is contained in P for all $i \in \mathbb{Z}_+$. Because of that, the endpoint moving bijection $M_i = M_{f(t)f(t_i)} : \mathcal{L}_{sf(t)} \to \mathcal{L}_{sf(t_i)}$ of Theorem 3.19 is defined, and for all $\gamma \in \mathcal{L}_{sf(t)}$ it holds that

$$\Big||M_i(\gamma)| - |\gamma|\Big| \leq ||f(t_i) - f(t)|| < 1/i. \tag{5}$$

Therefore $|M_i(\gamma)| \to |\gamma|$ when $i \to \infty$. Because L is continuous, we also know that $L(t_i) \to L(t)$ when $i \to \infty$. Because of these convergences, if $\gamma \in \mathcal{L}_{sf(t)}$ and $|\gamma| < L(t)$, then for all sufficiently large $i \in \mathbb{Z}_+$ it holds that

$|M_i(\gamma)| < L(t_i)$, and because there are only finite number of such γ, we get that there exists $i \in \mathbb{Z}_+$ such that M_i maps $\mathcal{L}_{sf(t),[0,L(t))}$ to $\mathcal{L}_{sf(t_i),[0,L(t_i))}$. Because M_i is injective, this yields that

$$|\mathcal{L}_{sf(t),[0,L(t))}| \leq |\mathcal{L}_{sf(t_i),[0,L(t_i))}| < k$$

where the last inequality follows from the fact that $f(t_i) \in T_{\leq k}(a, A; b, B)$.

By the continuity of L and the inequality (5) there exists $n \in \mathbb{Z}_+$ such that if $i \geq n$, $\gamma \in \mathcal{L}_{sf(t)}$ and $|\gamma| \geq L(t) + 1$, then $|M_i(\gamma)| > L(t_i)$. Furthermore, if $\gamma \in \mathcal{L}_{sf(t)}$ and $L(t) < |\gamma| < L(t) + 1$, then for sufficiently large $i \in \mathbb{Z}_+$ it holds that $|M_i(\gamma)| > L(t_i)$. Because there are only finitely many paths in $\mathcal{L}_{sf(t)}$ with length below $L(t) + 1$, we know that there exists $i \in \mathbb{Z}_+$ such that M_i maps $\mathcal{L}_{sf(t),(L(t),\infty)}$ to $\mathcal{L}_{sf(t_i),(L(t_i),\infty)}$. This means that the inverse M_i^{-1} maps $\mathcal{L}_{sf(t_i),[0,L(t_i)]}$ to $\mathcal{L}_{sf(t),[0,L(t)]}$. Because M_i^{-1} is injective, this yields that

$$|\mathcal{L}_{sf(t),[0,L(t)]}| \geq |\mathcal{L}_{sf(t_i),[0,L(t_i)]}| > k$$

where the last inequality follows from the fact that $f(t_i) \in T_{\leq k}(a, A; b, B)$. Thus we have proven that $f(t) \in T_{\leq k}(a, A; b, B)$. □

Theorem 5.24. *Let $\{(a, A), (b, B)\} \in T_{\leq k}(\cdot, \cdot)$, $f = f_{\mathcal{B}(a,A;b,B)}$ be as defined in Theorem 5.19, and $I = f^{-1}T_{\leq k}(a, A; b, B)$ be the set proved to be compact in Theorem 5.23. For all boundary points $t \in \partial I$ it holds that $f(t)$ is a k-vertex.*

Proof. Let $t \in \partial I$. If $f(t) \in \partial P$, then we get that $f(t) \in V_{\leq k}$ directly by definition. Thus it suffices to prove that if $t \in \partial I$ and $f(t) \notin \partial P$, then $f(t) \in V_{\leq k}$. Because $t \in \partial I$, there exists a sequence $(t_i)_{i=1}^{\infty}$ such that $t_i \in \mathbb{R} \setminus I$ for all $i \in \mathbb{Z}_+$, and $t_i \to t$ when $i \to \infty$. By Theorem 5.23, $t \in I$ and thus there exists $\alpha \in \mathcal{L}_{sa,\{A\}}$ and $\beta \in \mathcal{L}_{sb,\{B\}}$ such that $\alpha[a, f(t)], \beta[b, f(t)] \in \mathcal{L}_{sf(t)}$. Now because $t_i \notin I$ for all $i \in \mathbb{Z}_+$, at least one of the following holds:

(a) For some $\alpha \in \mathcal{L}_{sa,\{A\}}$ such that $\alpha[a, f(t)] \in \mathcal{L}_{sf(t)}$, $\alpha[a, f(t_i)] \notin \mathcal{L}_{sf(t_i)}$.

(b) For some $\beta \in \mathcal{L}_{sa,\{A\}}$ such that $\beta[b, f(t)] \in \mathcal{L}_{sf(t)}$, $\beta[b, f(t_i)] \notin \mathcal{L}_{sf(t_i)}$.

(c) $|\mathcal{L}_{sf(t_i),[0,L(t_i))}| \geq k$ or $|\mathcal{L}_{sf(t_i),[0,L(t_i)]}| \leq k$, and (a) and (b) do not hold.

Here L is the continuous function $\mathbb{R} \to \mathbb{R}$ defined by

$$L(t) = A + ||f(t) - a|| = B + ||f(t) - b||.$$

Furthermore, one of the above cases holds for a infinitely many $i \in \mathbb{Z}_+$. Thus we may choose the sequence $(t_i)_{i=1}^{\infty}$ such that one of the cases holds for all $i \in \mathbb{Z}_+$. Let us prove the claim in cases depending on which one it is.

(a) Now for all $i \in \mathbb{Z}_+$ at least one of the following holds:

(i) $[a, f(t_i)]$ is not contained in P.

(ii) For some $\alpha \in \mathcal{L}_{sa,\{A\}}$ such that $\alpha[a, f(t)] \in \mathcal{L}_{sf(t)}$,

$$\angle(\alpha[a, f(t_i)], |\alpha|) < 180°.$$

Again, by limiting $(t_i)_{i=1}^{\infty}$ to an infinite subsequence, we may assume that one of the above cases holds for all $i \in \mathbb{Z}_+$, and prove the claim in cases.

(i) Now because $f(t)$ is an interior point of P, $f(t_i)$ is also an interior point of P for sufficiently large $i \in \mathbb{Z}_+$. Therefore $[a, f(t_i)]$ intersects an edge of ∂P that is not incident to a. Because ∂P consists of finitely many finite polygons, there exists a subsequence $(t'_i)_{i=1}^{\infty}$ of $(t_i)_{i=1}^{\infty}$ such that this intersecting edge is always the same, i.e. $[x, y]$ is an edge of ∂P such that $x \neq a \neq y$ and $[a, f(t'_i)]$ intersects $[x, y]$ for all $i \in \mathbb{Z}_+$. Let c_i be this intersection point. Because $[x, y]$ is closed and f is continuous, we get that $(c_i)_{i=1}^{\infty}$ converges to the intersection point c of $[a, f(t)]$ and $[x, y]$.

Because $\alpha[a, f(t)] \in \mathcal{L}_{sf(t)}$, $[a, f(t)]$ is contained in P. Because $[a, f(t)]$ intersects the edge $[x, y]$, it must do so at the endpoint of the edge. Thus $c \in \{x, y\}$, which means that $c \in V \setminus \{a\}$. Define $\gamma = \alpha[a, c]$ and $C = |\gamma|$. Now $\gamma[c, f(t)] \in \mathcal{L}_{sf(t)}$, and using the identity $C = A + ||c - a||$, the assumption $|B - A| < ||b - a||$ and the triangle inequality, we can bound

$$\begin{aligned}|C - B| &= |A + ||c - a|| - B| \\ &\leq |A - B| + ||c - a|| \\ &< ||a - b|| + ||c - a|| \\ &\leq ||c - b||.\end{aligned}$$

Thus $f(t) \in T_{\leq k}(c, C; b, B)$ and $c \neq a$, which means that $f(t)$ is in an intersection of two distinct k-bisectors. This yields the claim that $f(t) \in V_{\leq k}$.

(ii) By Theorem 3.10, $\mathcal{L}_{sa,\{A\}}$ is finite, and thus there exists an infinite subsequence $(t'_i)_{i=1}^{\infty}$ of $(t_i)_{i=1}^{\infty}$ and $\alpha \in \mathcal{L}_{sa,\{A\}}$ such that $\alpha[a, f(t)] \in \mathcal{L}_{sf(t)}$ but $\angle(\alpha[a, f(t'_i)], |\alpha|) < 180°$ for all $i \in \mathbb{Z}_+$. Because f does not pass through a, the function $x \mapsto \angle(\alpha[a, f(x)], |\alpha|)$ is continuous in some neighborhood of t. Now we obtain the angle $\angle(\alpha[a, f(t)], |\alpha|)$ as a limit of the angles $\angle(\alpha[a, f(t_i)], |\alpha|)$ as $i \to \infty$. Because $\alpha[a, f(t)] \in \mathcal{L}_{sf(t)}$, $\angle(\alpha, [a, f(t)], |\alpha|) \geq 180°$, and because it is obtained as a limit of numbers less than 180°, it must be exactly 180°.

Because of this, there exists $c \in V \setminus \{a\}$ and $\gamma \in \mathcal{L}_{sc}$ that is a prefix of α such that $\gamma[c, f(t)] = \alpha[a, f(t)]$. Similarly to (i), we get that if $C = |\gamma|$, then $|C - B| < ||c - b||$ and $f(t) \in T_{\leq k}(c, C; b, B)$, which yields that $f(t) \in V_{\leq k}$.

(b) Analogous to (a).

(c) We will first prove that some $\xi \in \mathcal{L}_{sf(t),\{L(t)\}}$ does not have $[a, f(t)]$ or $[b, f(t)]$ as a suffix. Assume the contrary: all the paths of $\mathcal{L}_{sf(t),\{L(t)\}}$ have either $[a, f(t)]$ or $[b, f(t)]$ as a suffix. Because $f(t)$ is an interior point of P, for sufficiently large $i \in \mathbb{Z}_+$ the segment $[f(t), f(t_i)]$ is contained in P, and therefore the endpoint moving bijection

$$M_i = M_{f(t)f(t_i)} : \mathcal{L}_{sf(t)} \to \mathcal{L}_{sf(t_i)}$$

of Theorem 3.19 is defined. Because (a) does not hold, we know that for all $i \in \mathbb{Z}_+$ and $\alpha \in \mathcal{L}_{sa,\{A\}}$ such that $\alpha[a, f(t)] \in \mathcal{L}_{sf(t)}$ it holds that $\alpha[a, f(t_i)] \in \mathcal{L}_{sf(t_i)}$. Now if $M_i(\alpha[a, f(t)]) \neq \alpha[a, f(t_i)]$, then $\alpha[a, f(t)]$ is not homotopic to $\alpha[a, f(t_i)]$. This means that the interior of the triangle with vertices a, $f(t)$ and $f(t_i)$ contains points not in P even though the edges $[a, f(t)]$ and $[a, f(t_i)]$ are contained in P, and it is easy to see that this can hold for only finite number of $i \in \mathbb{Z}_+$. Thus for all sufficiently large $i \in \mathbb{Z}_+$, if $\alpha \in \mathcal{L}_{sa,\{A\}}$ and $\alpha[a, f(t)] \in \mathcal{L}_{sf(t)}$ then $M_i(\alpha[a, f(t)]) = \alpha[a, f(t_i)]$, and by a similar reasoning from the fact that (b) does not hold we know that also if $\beta \in \mathcal{L}_{sb,\{B\}}$ and $\beta[b, f(t)] \in \mathcal{L}_{sf(t)}$ then $M_i(\beta[b, f(t)]) = \beta[b, f(t_i)]$. By combining this with our counterassumption that all paths of $\mathcal{L}_{sf(t),\{L(t)\}}$ can be written either as $\alpha[a, f(t)]$ for some $\alpha \in \mathcal{L}_{sa,\{A\}}$ or $\beta[b, f(t)]$ for some $\beta \in \mathcal{L}_{sb,\{B\}}$, we get that M_i maps all paths with length $L(t)$ to paths with length $L(t_i)$.

Let $\delta > 0$ be such that for all $\gamma \in \mathcal{L}_{sf(t)}$ with $|\gamma| \neq L(t)$, $\delta < \Big||\gamma| - L(t)\Big|$. This kind of δ exists due to Theorem 3.10. Because

$$\Big||M_i(\gamma)| - |\gamma|\Big| \leq |f(t_i) - f(t)|$$

for all $\gamma \in \mathcal{L}_{sf(t)}$, for sufficiently large $i \in \mathbb{Z}_+$ independent of γ, $\Big||M_i(\gamma)| - |\gamma|\Big| < \delta/2$ for all $\gamma \in \mathcal{L}_{sf(t)}$. Because $L(t_i) \to L(t)$ when $i \to \infty$, for sufficiently large $i \in \mathbb{Z}_+$, $|L(t_i) - L(t)| < \delta/2$. By combining these two inequalities using the triangle inequality we get that there exists $i \in \mathbb{Z}_+$ such that for all $\gamma \in \mathcal{L}_{sf(t)}$ for which $|\gamma| \neq L(t)$,

$$|M_i(\gamma)| - L(t_i) \geq |\gamma| - \Big||M_i(\gamma)| - |\gamma|\Big| - L(t) - |L(t_i) - L(t)| > |\gamma| - L(t) - \delta,$$

and

$$|M_i(\gamma)| - L(t_i) \leq |\gamma| + \Big||M_i(\gamma)| - |\gamma|\Big| - L(t) + |L(t_i) - L(t)| < |\gamma| - L(t) + \delta.$$

Because $\Big||\gamma| - L(t)\Big| > \delta$, this means that the ordering of $|M_i(\gamma)|$ and $L(t_i)$ is the same as the ordering of $|\gamma|$ and $L(t)$. Now M_i is a bijection that maps every path with length equal to/less than/greater than $L(t)$ to a path of length equal to/less than/greater than $L(t_i)$, respectively. Because $t \in I$, it holds that $|\mathcal{L}_{sf(t),[0,L(t))}| < k < |\mathcal{L}_{sf(t),[0,L(t)]}|$, and since M_i preserves the ordering of paths of length $L(t)$ to other paths, $|\mathcal{L}_{sf(t_i),[0,L(t_i))}| < k < |\mathcal{L}_{sf(t_i),[0,L(t_i)]}|$, which is a contradiction with the assumption of case (c).

Now we have proven that there exists $\xi \in \mathcal{L}_{sf(t),\{L(t)\}}$ that does not have $[a, f(t)]$ or $[b, f(t)]$ as a suffix. Because ξ has length equal to some other locally shortest path, by Theorem 2.12 we know that $\xi \neq [s, f(t)]$ and therefore there exists $c \in V \setminus \{a, b\}$ and $\gamma \in \mathcal{L}_{sc}$ such that $\xi = \gamma[c, f(t)]$. Now if we set $C = |\gamma|$, then $f(t) \in T_{\leq k}(c, C; b, B)$ which means that $f(t)$ is in an intersection of two distinct k-bisectors. This yields the claim $f(t) \in V_{\leq k}$.

□

Lemma 5.25. *For all $k \in \mathbb{Z}_+$ and $x \in P$, denote the set of lengths of the k shortest paths in $\mathcal{L}_{sx}$ by $\mathcal{L}_{sx}(1 \dots k)$:*

$$\mathcal{L}_{sx}(1 \dots k) = \{\mathcal{L}_{sx}(1), \mathcal{L}_{sx}(2), \dots, \mathcal{L}_{sx}(k)\}.$$

- *If $a, b \in P$, $\alpha \in \mathcal{L}_{sa}$ and $\beta \in \mathcal{L}_{sb}$ such that α is a prefix of β, then $|\beta| \in \mathcal{L}_{sb}(1 \dots k)$ implies that $|\alpha| \in \mathcal{L}_{sa}(1 \dots k)$.*
- *If $\{(a, A), (b, B)\} \in T_{\leq k}(\cdot, \cdot)$ such that $T_{\leq k}(a, A; b, B) \neq \emptyset$, then it holds that $A \in \mathcal{L}_{sa}(1 \dots k)$ and $B \in \mathcal{L}_{sb}(1 \dots k)$.*

Proof.

- Choose $\gamma \in \mathcal{L}_{ab}$ such that $\alpha\gamma = \beta$. Define $f : \mathcal{L}_{sa} \to \mathcal{L}_{sb}$ such that $f(\xi)$ is the shortcut of $\xi\gamma$. Now f is an injection, because if $\xi, \phi \in \mathcal{L}_{sa}$ are distinct paths, then $\xi\gamma$ and $\phi\gamma$ are not homotopic, which yields that their shortcuts are distinct. From Theorem 3.18 we get that for all $\xi \in \mathcal{L}_{sa}$ it holds that $|f(\xi)| \leq |\xi| + |\gamma|$.

  Assume that $|\alpha| \notin \mathcal{L}_{sa}(1 \dots k)$, i.e. $|\alpha| > \mathcal{L}_{sa}(k)$. By the definition of $\mathcal{L}_{sa}(1 \dots k)$, there exist k distinct locally shortest paths $\alpha_1, \alpha_2, \dots, \alpha_k \in \mathcal{L}_{sa}$ with lengths at most $\mathcal{L}_{sa}(k)$. Because f is an injection, we get that $f(\alpha_1), f(\alpha_2), \dots, f(\alpha_k)$ are k distinct paths of $\mathcal{L}_{sb}$ with lengths at most $\mathcal{L}_{sa}(k)+|\gamma|$. Because $|\beta| = |\alpha|+|\gamma| > \mathcal{L}_{sa}(k)+|\gamma| \geq |\alpha_i|$ for all $i \in \{1, 2, \dots, k\}$, we have now proven that $|\beta| \notin \mathcal{L}_{sb}(1 \dots k)$. Now we have proven the contraposition of the implication of the claim, which means that the claim holds.

- If $x \in T_{\leq k}(a, A; b, B)$, then by definition, there exists $\alpha \in \mathcal{L}_{sa,\{A\}}$ such that $\alpha[a, x] \in \mathcal{L}_{sx,\{L\}}$ where $L = \mathcal{L}_{sx}(k)$. Now by the previous point we get that because $|\alpha[a, x]| \in \mathcal{L}_{sx}(1 \dots k)$, $A = |\alpha| \in \mathcal{L}_{sa}(1 \dots k)$. We get that $B = |\alpha| \in \mathcal{L}_{sb}(1 \dots k)$ analogously.

$\square$

Lemma 5.26. *Let $a, b \in V$ and $A, B \geq 0$ such that $|B - A| < ||b - a||$. Let S be a line in $\mathbb{R}$. Now if $|S \cap \mathcal{B}(a, A; b, B)| > 2$, then $A = B$ and S is the orthogonal bisector line of the segment ab.*

Proof. Consider the case where $a = (-1, 0)$ and $b = (1, 0)$. The rest of the cases follow by rotation, scaling and translation. In the proof of Theorem 5.19 we showed that the points of $\mathcal{B}(a, A; b, B)$ are exactly the points $(x, y) \in \mathbb{R}^2$ satisfying $x = g(y)$ where $g : \mathbb{R} \to \mathbb{R}$ is defined as follows:

$$g(y) = \frac{B - A}{2} \sqrt{\frac{4(y^2 + 1) - (B - A)^2}{4 - (B - A)^2}}.$$

Now

$$\frac{dg}{dy}(y) = \frac{2(B - A)y}{4 - (B - A)^2} \sqrt{\frac{4 - (B - A)^2}{4(y^2 + 1) - (B - A)^2}}$$

Now if $A < B$, then $\frac{dg}{dy}$ is strictly increasing, because $0 < B - A < 2$ and

$$\frac{\sqrt{4(y^2 + 1) - (B - A)^2}}{y} = \sqrt{4 + \frac{4 - (B - A)^2}{y^2}}$$

is strictly decreasing. Similarly we get that if $A > B$, then $\frac{dg}{dy}$ is strictly decreasing.

Clearly $\mathcal{B}(a, A; b, B)$ intersects any line $y = D$ exactly once. Consider a line S defined by $x = Cy + D$ for some $C, D \in \mathbb{R}$. Now if $(Cy + D, y)$ is an intersection point of $\mathcal{B}(a, A; b, B)$ and S, then $h(y) = g(y) - Cy - D$ is zero. If $A \neq B$, then $\frac{dg}{dy}$ is strictly monotonic, which means that $\frac{dh}{dy}$ is also strictly monotonic. Thus $\frac{dh}{dy}(y) = 0$ for at most one value y, and h is strictly monotonic everywhere else, which yields that $h(y)$ can be zero for at most 2 values y, and therefore $|S \cap \mathcal{B}(a, A; b, B)| \leq 2$. Otherwise $A = B$ and $g(y) = 0$, which means that $\mathcal{B}(a, A; b, B)$ is the line $y = 0$, that is, the orthogonal bisector line of the segment ab. $\square$

Theorem 5.27. *Each k-bisector contains $O(k|V|)$ k-vertices.*

Proof. Let $\{(a, A), (b, B)\} \in T_{\leq k}(\cdot, \cdot)$. Let f be the injection $\mathbb{R} \to \mathbb{R}^2$ that has image $\mathcal{B}(a, A; b, B)$ from Theorem 5.19. In Definition 5.22, the k-vertices were defined in two cases. Let us prove for both cases separately that there are $O(k|V|)$ vertices of that case in $T_{\leq k}(a, A; b, B)$.

- If $f(t)$ is a k-vertex as defined in the first case of Definition 5.22 for some $t \in \mathbb{R}$, then there exists $\{(c, C), (d, D)\} \in T_{\leq k}(\cdot, \cdot) \setminus \{\{(a, A), (b, B)\}\}$ such that $f(t) \in T_{\leq k}(c, C; d, D)$. Because the pair $\{(c, C), (d, D)\}$ is unordered, we may assume that $(c, C) \notin \{(a, A), (b, B)\}$. Now by definition, $A + ||f(t) - a||$, $B + ||f(t) - b||$ and $C + ||f(t) - c||$ are all equal to $\mathcal{L}_{sf(t)}(k)$, and therefore they are all equal to each other. If $c = a$, then because $C \neq A$, the equation $A + ||f(t) - a|| = C + ||f(t) - c||$ cannot hold, which is a contradiction. Similarly we get that $c = b$ yields a contradiction. Thus a, b and c are distinct vertices. Define for all $(c, C) \in V \times \mathbb{R}$ the set $V(c, C)$ by

  $$V(c, C) = \{f(t) \mid t \in \mathbb{R}, ||f(t) - c|| + C = ||f(t) - a|| + A\}.$$

  By Lemma 5.25 we know that if $C \notin \mathcal{L}_{sc}(1 \ldots k)$, then $V(c, C)$ is empty. Thus the set of k-vertices that satisfy the first case of Definition 5.22 and are contained in $T_{\leq k}(a, A; b, B)$ is a subset of

  $$\bigcup_{c \in V \setminus \{a,b\}, C \in \mathcal{L}_{sc}(1 \ldots k)} V(c, C).$$

  The size of the union is $O(k|V|)$, because for all $c \in V$ the set $\mathcal{L}_{sc}(1 \ldots k)$ has size at most k. To prove the claim, it suffices to prove that the sizes of the sets $V(c, C)$ are bounded by a constant.

  Let $c \in V \setminus \{a, b\}$ and $C \geq 0$. Let $g(t) = ||f(t) - c|| - ||f(t) - a|| + C - A$. Because for all $t \in \mathbb{R}$ it holds that $||f(t) - a|| + A = ||f(t) - b|| + B$, we get an alternative form $g(t) = ||f(t) - c|| - ||f(t) - b|| + C - B$. Now $f(t) \in V(c, C)$ if $g(t) = 0$, and thus we need to bound the number of zeros of g. Denote $c = (u, v)$. Let us compute the derivative of g:

  $$\frac{dg}{dt}(t) = \nabla f(t) \cdot (d_c - d_a) = \nabla f(t) \cdot (d_c - d_b).$$

  Here ∇ is the gradient operator, $\cdot$ is the dot product operator and d_v is equal to $\frac{f(t)-v}{||f(t)-v||}$. Note that the derivative is defined if $f(t) \neq c$ (we know that f does not pass through a or b because $|B - A| < ||b - a||$). Because f is injective, $\nabla f(t) \neq (0, 0)$.

  Now if $\frac{dg}{dt}(t) = 0$, then $d_c - d_a$ and $d_c - d_b$ are orthogonal to $\nabla f(t)$, and because d_a, d_b and d_c are unit vectors, we know that at least two of them are equal. If $d_a = d_b$, then a and b are in the same direction from $f(t)$, which means that $f(t)$ is on line ab and $|A - B| = ||a - b||$, which is a contradiction because we assumed that $\{(a, A), (b, B)\} \in T_{\leq k}(\cdot, \cdot)$. Thus $f(t)$ is on at least one of the lines ac and bc.

  Because the vertices a, b and c are not collinear, ac and bc are not equal to the bisector of the segment ab, and thus by Lemma 5.26 we get that $\frac{dg}{dt}(t)$ can be zero for at most four values of t. Because g is continuously differentiable in points where $f(t) \neq c$, and $f(t) = c$ can hold for at most one $t \in \mathbb{R}$, this yields that the equation $g(t) = 0$ has at most 6 solutions, and thus $|V(c, C)| \leq 6$.

- To prove that $T_{\leq k}(a, A; b, B)$ contains $O(k|V|)$ k-vertices satisfying the second case of Definition 5.22, we will prove a stronger claim: $T_{\leq k}(a, A; b, B)$ contains $O(|V|)$ points in ∂P. The boundary ∂P consists of edges $[u_1, v_1], [u_2, v_2], \ldots, [u_n, v_n]$ where $n = O(|V|)$. We will prove the claim by proving that

$$|T_{\leq k}(a, A; b, B) \cap [u_i, v_i]| \leq 2$$

for all $i \in \{1, 2, \ldots, n\}$.

Let $i \in \{1, 2, \ldots, n\}$. If $[u_i, v_i]$ is not a part of the orthogonal bisector line of segment ab, then we get $|T_{\leq k}(a, A; b, B) \cap [u_i, v_i]| \leq 2$ directly from Lemma 5.26, because $T_{\leq k}(a, A; b, B) \subset \mathcal{B}(a, A; b, B)$. Assume that $[u_i, v_i]$ is a part of the orthogonal bisector line of segment ab. Now we know that a and b are on different sides of the line $u_i v_i$, and thus for all $0 < t < ||v_i - u_i||$, one of the segments $[a, [u_i, v_i](t)]$ and $[b, [u_i, v_i](t)]$ is not contained in P, yielding that $[u_i, v_i](t) \notin T_{\leq k}(a, A; b, B)$. Therefore $T_{\leq k}(a, A; b, B) \cap [u_i, v_i] \subset \{u_i, v_i\}$ and $|T_{\leq k}(a, A; b, B) \cap [u_i, v_i]| \leq 2$.

□

The k-bisectors are possibly disconnected and might intersect each other. The above theorem guarantees that each k-bisector contains only a finite number of k-vertices, and thus we can split the k-bisectors into a finite number of *k-wall segments*, that is, the parts of k-bisectors between adjacent k-vertices. This way we get simpler objects than k-bisectors to work with: the k-wall segments are connected curves between k-vertices that do not intersect each other.

Definition 5.28. Let $k \in \mathbb{Z}_+$, $\{(a, A), (b, B)\} \in T_{\leq k}(\cdot, \cdot)$. Let $f = f_{\mathcal{B}(a,A;b,B)}$ be as defined in Theorem 5.19. Let $t_1 < t_2 < \ldots < t_n$ be the values $t \in \mathbb{R}$ such that $f(t) \in T_{\leq k}(a, A; b, B)$ and $f(t) \in V_{\leq k}$. We say that $f[t_i, t_{i+1}]$ is a *k-wall segment* if $f[t_i, t_{i+1}] \subset T_{\leq k}(a, A; b, B)$ and $i \in \{1, 2, \ldots, n-1\}$. Denote the set of k-wall segments by $E_{\leq k}$. As a special case, define that the set of 0-wall segments $E_{\leq 0}$ is an empty set.

Lemma 5.29. *For all $k \in \mathbb{Z}_+$, $T_{\leq k}$ does not contain isolated points.*

Proof. Assume the contrary: $x \in T_{\leq k}$ is an isolated point of $T_{\leq k}$. Let $L = \mathcal{L}_{sx}(k)$. Now by Theorem 5.17 we know that $\mathcal{L}_{sx,[0,L)} < k < \mathcal{L}_{sx,[0,L]}$. Let R be an equivalence relation in $\mathcal{L}_{sx,\{L\}}$ such that paths that come from the same direction are equivalent, i.e. $\alpha R \beta$ if $\mathrm{LCS}(\alpha, \beta) > 0$. By Theorem 5.17 we know that there are at least two equivalence classes, i.e. $|\mathcal{L}_{sx,\{L\}}/R| \geq 2$.

Let us use $r_x > 0$ and e_γ^{-1} from Theorem 3.13. All the paths in any equivalence class $S \in \mathcal{L}_{sx,\{L\}}/R$ are totally ordered by $\preceq$, and by Lemma 5.4 we know the bijective map $\gamma \mapsto e_\gamma^{-1}(y)$ preserves the $\preceq$-ordering in S for all $y \in B_P(x, r_x)$.

By Lemma 5.5 we can choose $0 < r < r_x$ small enough such that $\gamma \mapsto e_\gamma^{-1}(y)$ preserves the ordering by length between all the paths in $\mathcal{L}_{sx,\{L\}}$ and all the other paths of $\mathcal{L}_{sx}$ for all $y \in B_P(x, r)$. By Lemma 3.7 we may choose r such that $B_P(x, r)$ is a disk sector.

Let $S \in \mathcal{L}_{sx,\{L\}}/R$. As all the paths of S come from the same direction, $\gamma(|\gamma| - r/2)$ is the same point for all $\gamma \in S$. Denote this point by x_S. Now $[x_S, x]$ is a common suffix for all paths of S. For all $\gamma \in S$, the path $e_\gamma^{-1}(x_S)$ is obtained just by removing the common suffix $[x_S, x]$, and thus $|e_\gamma^{-1}(x_S)| = L - r/2$. Furthermore, if $\gamma \in \mathcal{L}_{sx,\{L\}} \setminus S$, then by the explicit formula for e_γ^{-1} of Theorem 3.13 we know that $e_\gamma^{-1}(x_S) = \gamma_{[0,t]}[\gamma(t), x_S]$, where $t \in [0, |\gamma|]$ such that $[\gamma(t), x]$ is a suffix of γ, and now

$$e_\gamma^{-1}(x_S) = L - ||x - \gamma(t)|| + ||x_S - \gamma(t)|| < L - ||x_S - x|| = L - r/2,$$

where the inequality follows from the triangle inequality, and the inequality is strict because $\gamma \notin S$ and thus x_S is not between $\gamma(t)$ and x. Therefore, when moving the endpoints to x_S using the map $\gamma \mapsto e_\gamma^{-1}(x_S)$, the paths of S are strictly shortest among the paths of $\mathcal{L}_{sx,\{L\}}$.

Let $B \in \mathcal{L}_{sx,\{L\}}/R$. From the length order preserving property and the fact that $\mathcal{L}_{sx,[0,L]} > k$ we know that there exists $A \in \mathcal{L}_{sx,\{L\}}/R$ such that $A \neq B$ and the set $\{\gamma \in A \mid e_\gamma^{-1}(x_B) \notin p_{\leq k}(x_B)\}$ is nonempty. Because the set is totally ordered by $\preceq$, it contains a minimum element α. When moving the endpoint from x_B but x_A, the paths of A preserve their ordering but become the strictly shortest among the paths of $\mathcal{L}_{sx,\{L\}}$, and because they switch places at least with the paths of B, by the minimality of α we get that $e_\alpha^{-1}(x_A) \in p_{\leq k}(x_A)$.

Let ξ be a path in $B_P(x, r) \setminus \{x\}$ from x_A to x_B. We just proved that $e_\alpha^{-1}(\xi(t)) \in p_{\leq k}(\xi(t))$ holds if $t = 0$ but does not hold if $t = |\xi|$. Define $f, g : [0, |\xi|] \to \mathbb{Z}$ by

$$\begin{aligned} f(t) &= \left|\left\{\gamma \in A \setminus \{\alpha\} \;\middle|\; \gamma \preceq \alpha\right\}\right| + \left|\left\{\gamma \in \mathcal{L}_{sx} \setminus A \;\middle|\; |e_\gamma^{-1}(\xi(t))| < |e_\alpha^{-1}(\xi(t))|\right\}\right|, \\ g(t) &= \left|\left\{\gamma \in A \;\middle|\; \gamma \preceq \alpha\right\}\right| + \left|\left\{\gamma \in \mathcal{L}_{sx} \setminus A \;\middle|\; |e_\gamma^{-1}(\xi(t))| \leq |e_\alpha^{-1}(\xi(t))|\right\}\right|. \end{aligned}$$

Note that the first term of both definitions is a constant, and by the consistency of the $\preceq$-ordering of paths in A we know that $|\{\gamma \in A \mid \gamma \preceq \alpha\}|$ is the same as $|\{\gamma \in A \mid e_\gamma^{-1}(\xi(t)) \preceq e_\alpha^{-1}(\xi(t))\}|$. Now it suffices to prove that for some $t \in [0, |\xi|]$ it holds that $f(t) < k < g(t)$, because then we get that contradiction $\xi(t) \in T_{\leq k}$ from Theorem 5.17: we know that $\mathcal{L}_{sx}(k) = |e_\alpha^{-1}(\xi(t))|$ and because $g(t) - f(t) \geq 2$, there exists $\beta \in \mathcal{L}_{sx} \setminus B$

such that $|e_\alpha^{-1}(\xi(t))| = |e_\beta^{-1}(\xi(t))|$ and as $\beta \notin A$, the paths do not have nontrivial common suffix.

Because the lengths of the paths given by $\gamma \mapsto e_\gamma^{-1}(\xi(t))$ change continuously as a function of t, we get that f is right-continuous and g is left-continuous, i.e. $f(t) = \lim_{s\to t+} f(s)$ and $g(t) = \lim_{s\to t-} f(s)$ for all $0 < t < |\xi|$. Now because $f(0) > k$ and $f(|\xi|) < k$, we know that there exists a minimum t such that $f(t) < k$, and because $g(s) > f(s)$ for all $s \in [0, |\xi|]$, we then get that $g(t) \lim_{s\to t-} g(t) > k$. Thus $f(t) < k < g(t)$. $\square$

Theorem 5.30. *For all $k \in \{0, 1, 2, \ldots\}$, the k-wall segments fill the set of k-walls exactly, i.e.*

$$\bigcup E_{\leq k} = T_{\leq k}.$$

Proof. The case $k = 0$ clearly holds because $T_{\leq 0} = \emptyset$. Assume that $k \in \mathbb{Z}_+$.

Because all the sets of $E_{\leq k}$ are subsets of k-bisectors, by Theorem 5.21 we know that $\bigcup E_{\leq k} \subset T_{\leq k}$. Thus it suffices to prove that $T_{\leq k} \subset \bigcup E_{\leq k}$. Assume that $x \in T_{\leq k}$. Theorem 5.21 yields that there exists $\{(a, A), (b, B)\} \in T_{\leq k}(\cdot, \cdot)$ such that $x \in T_{\leq k}(a, A; b, B)$. Let $f = f_{\mathcal{B}(a,A;b,B)}$ and I be as in Theorem 5.23. Now $x = f(t)$ for some $t \in I$. Because I is compact, the component containing t in I is $[u, v]$ for some $u \leq t$ and $v \geq t$. Let us continue the proof in cases depending on whether $x \in V_{\leq k}$.

- Assume that $x \notin V_{\leq k}$. Let $u' = \max\{s \leq t \mid f(s) \in V_{\leq k}\}$ and $v' = \min\{s \geq t \mid f(s) \in V_{\leq k}\}$ (the maximum and minimum exist due to Theorem 5.27). Now $u \leq u' < t < v' \leq v$, because Theorem 5.24 yields that $f(u), f(v) \in V_{\leq k}$ and we assumed that $t \notin V_{\leq k}$. Now $f[u', v']$ is a k-wall segment containing $x = f(t)$, and thus $x \in \bigcup E_{\leq k}$.
- Assume that $x \in V_{\leq k}$. If $u < t$, by setting $u' = \max\{s < t \mid f(s) \in V_{\leq k}\}$ we get that $u \leq u' < t$ and thus $f[u', v']$ is a k-wall segment containing x, proving that $x \in \bigcup E_{\leq k}$. We get that also if $v > t$. If we consider all the possible choices $\{(a, A), (b, B)\} \in T_{\leq k}(\cdot, \cdot)$ such that $x \in T_{\leq k}(a, A; b, B)$, then the only case remaining is where x is an isolated point in $T_{\leq k}(a, A; b, B)$ for all $\{(a, A), (b, B)\} \in T_{\leq k}(\cdot, \cdot)$ such that $x \in T_{\leq k}(a, A; b, B)$.

  Theorem 5.23 yields that all k-bisectors are compact and by Lemma 5.25 we know that there are only finitely many nonempty k-bisectors. Therefore x is an isolated point of $T_{\leq k}$, which combined with Lemma 5.29 yields a contradiction.

$\square$

Theorem 5.31. *The number of k-wall segments in $E_{\leq k}$ is $O(k^3|V|^3)$.*

Proof. From Lemma 5.25 we know that each element of $E_{\leq k}$ is a subset of $T_{\leq k}(a, A; b, B)$ for some $\{(a, A), (b, B)\} \in T_{\leq k}(\cdot, \cdot)$ such that $A \in \mathcal{L}_{sa}(1 \dots k)$ and $B \in \mathcal{L}_{sb}(1 \dots k)$. Because $|\mathcal{L}_{sv}(1 \dots k)| \leq k$ for all $v \in V$, there are $O(k^2|V|^2)$ such pairs $\{(a, A), (b, B)\}$. By Theorem 5.27 we know that each k-bisector contains at most $O(k|V|)$ k-wall segments. By multiplying $O(k^2|V|^2)$ by $O(k|V|)$ we get the total $O(kv \cdot k^2|V|^2) = O(k^3|V|^3)$ □

5.5 Querying kth shortest paths

We proved in Theorem 5.9 that for $x \in M_k$, we can find the k-path from s to x by finding the component M_k containing x. In this subsection, we will collect the results of the previous subsections to prove that the k-SPM M_k can be represented by a data structure from which we can efficiently query kth shortest paths from s to any given point. The idea is to use standard point location query data structures that support finding the component of a point in a set if the complement of the set is consists of well-behaved curves that do not internally intersect each other. We are quite close to that already, as we know that $M_k = P_k \setminus W_k = P \setminus (T_k \cup W_k)$, and we have proved in the previous subsections that $T_k = T_{\leq k} \cup T_{\leq k-1}$ and W_k consists of a finite number of hyperbolic and line segments.

However, even though the sets W_k and $T_{\leq k} \cup T_{\leq k-1}$ separate the components of M_k in the domain P, they generally do not do that in the whole space $\mathbb{R}^2$: we need to add the boundaries of P. For simplicity, let us cut off the points of ∂P from M_k and limit our consideration to an open subset M'_k.

Lemma 5.32. *For all $k \in \mathbb{Z}_+$, define*

$$M'_k = M_k \setminus \partial P = P \setminus (W_k \cup T_k \cup \partial P)$$

Now M'_k is an open subset of $\mathbb{R}$, and it has the same property as M_k in Theorem 5.9.

Proof. By Theorem 5.30 we know that $T_k = T_{\leq k} \cup T_{\leq k-1}$ consists of a finite number of k-wall segments, and as k-wall segments are compact, T_k is closed. By Theorem 5.12 we know that W_k consists of a finite number of line segments, and their endpoints are contained in $T_k \cup \partial P$. Thus $W_k \cup T_k \cup \partial P$ is closed.

We know that $P \setminus \partial P$ is open because it is the interior of P, and because $W_k \cup T_k \cup \partial P$ is closed, we get that $M_k = (P \setminus \partial P) \cap (\mathbb{R} \setminus (W_k \cup T_k \cup \partial P))$ is also open. Because M'_k is a subset of M_k, its components are subsets of components of M_k and thus it has the same property as M_k in Theorem 5.9. □

Even after cutting away ∂P from M_k, our method can still handle queries to almost every point in P, as shown by the following theorem.

Theorem 5.33. *For all $k \in \mathbb{Z}_+$, $P \setminus M'_k$ has measure zero. Consequently, almost every point in P is in M'_k.*

Proof. By Theorem 5.12 we know that W_k consists of finite number of line segments, and thus has measure zero. By Theorem 5.4 we know that $T_k = T_{\leq k} \cup T_{\leq k-1}$, and Theorems 5.30 and 5.31 yield that both sets consist of finite number of hyperbolic segments, which yields that T_k has measure zero. We defined ∂P as a finite union of polygons, and thus it has measure zero. Now $P \setminus M'_k = W_k \cup T_k \cup \partial P$ is a finite union of sets of measure zero, and thus has measure zero. $\square$

Theorem 5.34. *For all $k \in \mathbb{Z}_+$ it holds that*

$$\partial M'_k = P \setminus M'_k = W_k \cup T_k \cup \partial P,$$

where ∂ denotes the boundary operation in space $\mathbb{R}^2$.

Proof. By Lemma 5.32 we know that M'_k is open, and thus $\partial M'_k$ and M'_k are disjoint. Furthermore, because P is closed and $M'_k \subset P$ we know that $\partial M'_k \subset P \setminus M'_k$. Now it suffices to prove that $P \setminus M'_k \subset \partial M'_k$, because by Lemma 5.32 we know that $P \setminus M'_k = W_k \cup T_k \cup \partial P$.

Assume the contrary: there exists $x \in P \setminus M'_k$ such that $x \notin \partial M'_k$. Because $x \notin M'_k$ and $x \notin \partial M'_k$, x is not in the closure of M'_k which means that there exists $r > 0$ such that $B_P(x, r) \subset P \setminus M'_k$. By Lemma 3.7 we know that $B_P(x, r)$ has nonzero measure, which yields that $P \setminus M'_k$ has nonzero measure. This is a contradiction with the result of Theorem 5.33. $\square$

Now from the above theorem we get that by splitting the plane $\mathbb{R}^2$ by $\partial M'_k = W_k \cup T_k \cup \partial P$ we get a subdivision of the plane such that all the components of M'_k are there.

Corollary 5.35. *Let $k \in \mathbb{Z}_+$. Define that $\partial M'_k = W_k \cup T_k \cup \partial P$ is the k*th separating set*, denoted by X_k. Define that C_1 is the set of components of $\mathbb{R}^2 \setminus X_k$, and C_2 is the set of components of M'_k. Now*

$$\{X \in C_1 \mid X \cap P \neq \emptyset\} = C_2.$$

To be able to analyze the complexity and query time of the point location data structure, let us see how many non-intersecting curves we have to use to assemble the kth separating set. For that we need the following compatibility result stating that when advancing from the set of $(k-1)$-walls to the set of k-walls, some points outside $T_{\leq k-1}$ might become k-vertices, and some $(k-1)$-vertices might fall outside $T_{\leq k}$, but no $(k-1)$-vertices become points of $T_{\leq k} \setminus V_{\leq k}$ and no points of $T_{\leq k-1} \setminus V_{\leq k-1}$ become k-vertices.

Lemma 5.36. *Let $k, k' \in \{0, 1, 2, \ldots\}$ such that $|k - k'| = 1$.*

- *If $x \in T_{\leq k'}(a, A; b, B)$ for some $\{(a, A), (b, B)\} \in T_{\leq k'}(\cdot, \cdot)$, then either $x \in T_{\leq k}(a, A; b, B)$ or $x \notin T_{\leq k}$.*

- *If $x \in V_{\leq k'}$, then $x \in V_{\leq k}$ or $x \notin T_{\leq k}$.*

Proof.

- If $\mathcal{L}_{sx}(k) < \mathcal{L}_{sx}(k+1)$, then $\mathcal{L}_{sx,[0,\mathcal{L}_{sx}(k)]}$ is a k-pathset, yielding the claim $x \notin T_{\leq k}$. Let us assume the other case: $\mathcal{L}_{sx}(k) = \mathcal{L}_{sx}(k+1)$. From the definition of $T_{\leq k'}(a, A; b, B)$ we know that if we set L to $A+||x-a|| = B+||x-b||$, then $|\mathcal{L}_{sx,[0,L)}| < k' < |\mathcal{L}_{sx,[0,L]}|$. Now because $\mathcal{L}_{sx}(k) = \mathcal{L}_{sx}(k+1)$, then if $k' = k-1$, the fact that $|\mathcal{L}_{sx,[0,L]}| \geq k$ implies that $|\mathcal{L}_{sx,[0,L]}| > k$, and if $k' = k+1$, then the fact that $|\mathcal{L}_{sx,[0,L)}| \leq k$ implies that $|\mathcal{L}_{sx,[0,L)}| < k$. Thus in both cases, we get that $|\mathcal{L}_{sx,[0,L)}| < k < |\mathcal{L}_{sx,[0,L]}|$, and therefore $x \in T_{\leq k}(a, A; b, B)$.

- Because $x \in V_{\leq k'}$, we get that x is either a boundary point of P on a k'-bisector or a common point of two k'-bisectors. By the previous part, we get that if $x \in T_{\leq k}$, then x is either a boundary point of P on a k-bisector or a common point of two k-bisectors. Thus $x \in V_{\leq k}$.

□

Theorem 5.37. *Let E be the set of boundary edges of the polygonal domain P. For any $k \in \mathbb{Z}_+$, let $\bar{W}_k$ be the set of the closures of the line segments of W_k (that is, the k-window segments with their endpoints). Now we get the following representation for the kth separating set:*

$$X_k = \bigcup(\bar{W}_k \cup E_{\leq k} \cup E_{\leq k-1} \cup E).$$

The set $\bar{W}_k \cup E_{\leq k} \cup E_{\leq k-1} \cup E$ consists of $O(k^3|V|^3)$ closed line segments and closed hyperbolic segments. See Figure 5.6 for an example. If two distinct segments of $\bar{W}_k \cup E_{\leq k} \cup E_{\leq k-1} \cup E$ intersect at x, then x is endpoint of one of the segments.

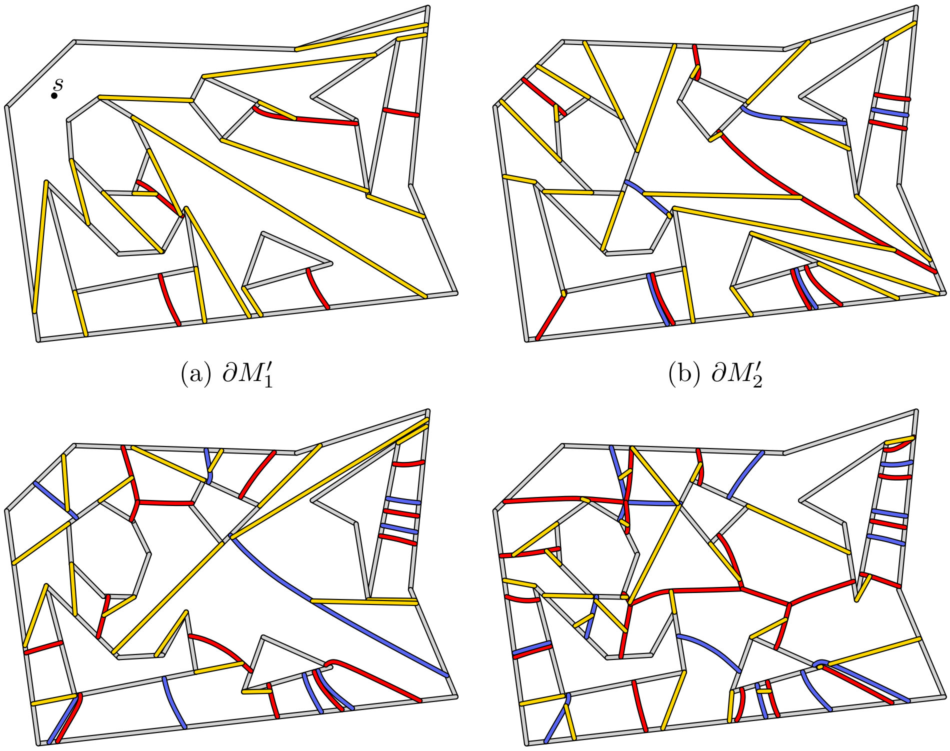
Proof. The claim that $X_k = \bigcup(\bar{W}_k \cup E_{\leq k} \cup E_{\leq k-1} \cup E)$ follows directly from the fact that it was defined as $W_k \cup T_k \cup \partial P = W_k \cup T_{\leq k} \cup T_{\leq k-1} \cup \partial P$, and all the endpoints of the segments in W_k are already in the set T_k, which means that their addition in $\bar{W}_k$ does not change the union.

We get the bound $O(k^3|V|^3)$ for the number of segments in the set $\bar{W}_k \cup E_{\leq k} \cup E_{\leq k-1} \cup E$ from the following:

- Theorem 5.12 yields that $|\bar{W}_k| = O(k|V|)$.

- Theorem 5.31 yields that $|E_{\leq k}|$ and $|E_{\leq k-1}|$ are both $O(k^3|V|^3)$.

- Because P is a polygonal domain, $|E| = |V|$.



(a) $\partial M_1'$ (b) $\partial M_2'$ (c) $\partial M_3'$ (d) $\partial M_4'$

Figure 5.6: The subfigures show the boundary of M_k' for $k \in \{1, 2, 3, 4\}$ with the same P and s as in Figure 5.4. As in Theorem 5.37, we see that the boundary of M_k' consists of the k-window segments (in yellow), k-wall segments (in red), $(k-1)$-wall segments (in blue) and the boundary segments of P (in white). These are either line segments or hyperbolic segments, and they do not intersect internally.

Let $A, B \in \bar{W}_k \cup E_{\leq k} \cup E_{\leq k-1} \cup E$. Let us prove that $A = B$ or any intersection point $x \in A \cap B$ is an endpoint of one of the segments A and B in cases:

- Case $A, B \in \bar{W}_k$: Follows directly from Theorem 5.12.
- Case $A, B \in E_{\leq k}$: x is a k-vertex, and thus x is an endpoint of both A and B.
- Case $A, B \in E_{\leq k-1}$: Analogous to the previous case.
- Case $A, B \in E$: Follows from the definition of a polygonal domain.
- Case $A \in \bar{W}_k$, $B \in E_{\leq i}$ for some $i \in \{k, k-1\}$: Because $x \in T_{\leq k} \cup T_{\leq k-1}$, $x \in T_k$. However, if $x \in W_k$ we get that $x \in P_k$, which means

that x must be an endpoint of A.

- Case $A \in \bar{W}_k$, $B \in E$: If x is not an endpoint of A, then we get that x is collinear with two distinct vertices $a, b \in V$ and x is on some edge on ab that is not $[a, b]$, which is a contradiction with the assumption that no three vertices are collinear.

- Case $A \in E_{\leq i}$, $B \in E$ for some $i \in \{k, k-1\}$: x is a k-vertex, and thus x is an endpoint of A.

- Case $A \in E_{\leq k-1}$, $B \in E_{\leq k}$: Let $\{(a, A), (b, B)\} \in T_{\leq k-1}(\cdot, \cdot)$ such that $A \subset T_{\leq k-1}(a, A; b, B)$. Now because $x \in T_{\leq k}$, by Lemma 5.36 we know that $x \in T_{\leq k}(a, A; b, B)$. If x is a k-vertex, then it is an endpoint of B. Assume that x is not a k-vertex. Now it holds that $A, B \subset \mathcal{B}(a, A; b, B)$, because if $B \not\subset T_{\leq k}(a, A; b, B)$, then B is a subset of some other k-bisector, which would yield that x is a k-vertex. We also know by Lemma 5.36 that x is not a $(k-1)$-vertex. Thus x is an internal point of a $(k-1)$-wall segment A and a k-wall segment B, and we want to prove that $A = B$.

  If $A \neq B$, then one of the endpoints of A is an internal point of B or vice versa. Any endpoint y of A is a $(k-1)$ vertex, and now Lemma 5.36 yields that either $y \notin T_{\leq k}$ or $y \in V_{\leq k}$, which means that y cannot be an internal point of B. A similar reasoning also proves that any endpoint of B cannot be an internal point of A. Thus $A = B$.

The rest of the cases follow by symmetry by swapping A and B. $\square$

As a corollary of the above theorem we get the space complexity of the kth shortest path map.

Corollary 5.38. *The kth shortest path map can be represented as a planar subdivision of complexity $O(k^3|V|^3)$ consisting of a set of line segments and hyperbolic segments that do not intersect internally.*

Now we have a set S of line segments and hyperbolic segments that do not intersect internally and make up the separating set X_k, and by Theorem 5.9 and Corollary 5.35 we can find the shortest path from s to any $x \in P$ by finding the component of $\mathbb{R}^2 \setminus X_k = \mathbb{R}^2 \setminus \bigcup S$ containing x. However, plainly storing S as a list does not support finding the component efficiently. We can fix this similarly to what was done in [10] for shortest path queries: by using standard point location query algorithms. We can build an $O(|S| \log |S|)$ space data structure in $O(|S| \log |S|)$ time to support queries in $O(\log |S|)$ time [2, 11]. The bound of Theorem 5.37 means that $|S| = O(k^3|V|^3)$. By Theorem 5.33, a component of M'_k is found for almost every $x \in P$. Let us formulate this result as a theorem.

Theorem 5.39. *For any $k \in \mathbb{Z}_+$, there exists a data structure with space complexity $O(k^3|V|^3(\log k + \log|V|))$ such that for almost every $x \in P$, we can query the kth shortest path from s to x in $O(\log k + \log|V|)$ time.*

Even though the data structure supports queries to only almost every $x \in P$, we can easily modify it to give reasonable answers for all $x \in P$, if we relax our strict definition of kth shortest paths to allow ties with paths that come from different directions: If $x \notin M_k'$, then we can just return the kth shortest path given by a neighboring component of M_k'.

References

[1] Alt, H. and Godau, M.: *Computing the Fréchet distance between two polygonal curves.* International Journal of Computational Geometry & Applications, 5(1–2):75–91, 1995.

[2] Edelsbrunner, H., Guibas, L. J., and Stolfi, J.: *Optimal point location in a monotone subdivision.* SIAM Journal on Computing, 15(2):317–340, 1986.

[3] Eppstein, D.: *Finding the k shortest paths.* SIAM Journal on Computing, 28(2):652–673, 1999.

[4] Eriksson-Bique, S., Hershberger, J., Polishchuk, V., Speckmann, B., Suri, S., Talvitie, T., Verbeek, K., and Yıldız, H.: *Geometric k shortest paths.* In *ACM-SIAM Symposium on Discrete Algorithms, San Diego, USA*, 2015.

[5] Fredman, M. L. and Tarjan, R. E.: *Fibonacci heaps and their uses in improved network optimization algorithms.* Journal of the ACM, 34(3):596–615, 1987.

[6] Ghosh, S. K. and Mount, D. M.: *An output-sensitive algorithm for computing visibility graphs.* SIAM Journal on Computing, 20(5):888–910, 1991.

[7] Hatcher, A.: *Algebraic topology.* Cambridge University Press, 2002.

[8] Hershberger, J., Polishchuk, V., Speckmann, B., and Talvitie, T.: *Geometric kth shortest paths: The applet.* In *23rd Video Review of Computational Geometry, Kyoto, Japan*, 2014. `http://www.computational-geometry.org/SoCG-videos/socg14video/ksp/index.html`.

[9] Hershberger, J and Snoeyink, J.: *Computing minimum length paths of a given homotopy class.* Computational Geometry, 4(2):63–97, 1994.

[10] Hershberger, J. and Suri, S.: *An optimal algorithm for Euclidean shortest paths in the plane.* SIAM Journal on Computing, 28(6):2215–2256, 1999.

[11] Kirkpatrick, D.: *Optimal search in planar subdivisions.* SIAM Journal on Computing, 12(1):28–35, 1983.

[12] Pocchiola, M. and Vegter, G.: *Topologically sweeping visibility complexes via pseudotriangulations.* Discrete & Computational Geometry, 16(4):419–453, 1996.